Learning and Instruction 66 (2020) 101271

Contents lists available at ScienceDirect

Learning and Instruction

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

Self-derivation through memory integration: A model for accumulation of semantic knowledge

Patricia J. Bauer[a,*,1], Alena G. Esposito[b,1], James J. Daly[a]

[a] Department of Psychology, Emory University, USA
[b] Department of Psychology, Clark University, USA

ARTICLE INFO

Keywords:
Learning
Memory
Memory integration
Semantic knowledge
Self-derivation

ABSTRACT

Semantic knowledge accumulates through explicit means and productive processes (e.g., analogy). These means work in concert when information explicitly acquired in separate episodes is integrated, and the integrated representation is used to self-derive new knowledge. We tested whether (a) self-derivation through memory integration extends beyond general information to science content, (b) self-derived information is retained, and (c) details of explicit learning episodes are retained. Testing was in second-grade classrooms (children 7–9 years). Children self-derived new knowledge; performance did not differ for general knowledge (Experiment 1) and science curriculum facts (Experiment 2). In Experiment 1, children retained self-derived knowledge over one week. In Experiment 2, children remembered details of the learning episodes that gave rise to self-derived knowledge; performance suggests that memory integration is dependent on explicit prompts. The findings support nomination of self-derivation through memory integration as a model for accumulation of semantic knowledge and inform the processes involved.

1. Introduction

Building a semantic knowledge base is a major task of development and education. Information accumulates through direct means, such as reading, listening to classroom lessons, interacting with museum exhibits, and so forth. Knowledge also expands through inferential processes including analogy, deduction, and induction (Gentner, 1983, 1989; Goswami, 2011). These means work in concert when information that is explicitly acquired in one learning episode is integrated with information explicitly acquired in a different episode, and through inferential processes operating over the newly integrated representation, new factual knowledge is generated or self-derived. Children ages 4–11 years (e.g., Bauer, Blue, Xu & Esposito, 2016a; Bauer & Larkina, 2017; Bauer & San Souci, 2010) and college students (e.g., Varga & Bauer, 2017a; 2017b) self-derive new factual knowledge through memory integration. They also retain self-derived knowledge over time (Varga & Bauer, 2013, 2017b; Varga, Stewart, & Bauer, 2016). Although self-derivation through integration has been examined in elementary classrooms (Esposito & Bauer, 2017, 2019), there have been no tests of whether classroom derived factual information is retained over a delay. Moreover, little is known about memory for the details of the episodes of direct learning that serve as the foundation for self-derivation. Accordingly, in Experiment 1, we examined retention of information self-derived in the classroom; the stimuli were drawn from prior laboratory studies. In Experiment 2, we examined memory for details of the explicit learning episodes upon which self-derivation depends; the stimuli were aligned with the children's science curriculum. Across experiments, we compared self-derivation of new factual knowledge across general and science domains.

Self-derivation through memory integration is one member of a broader class of inferential processes, including analogy, deduction, and induction (Gentner, 1983, 1989; Goswami, 2011). These processes are generally recognized to be major mechanisms of cognitive development (Bauer, 2012; Gentner, 1983, 1989; Goswami, 2011; Siegler, 1989). Self-derivation of new factual knowledge through memory integration occurs when information newly learned in one episode (e.g., *dolphins live in groups called* pods) is integrated with information newly learned in a separate, related episode (i.e., *dolphins talk by clicking and squeaking*). The integrated memory representation then supports productive extension of knowledge, such as the answer to the question "*How does a pod talk*?" Although the information that *pods talk by clicking and squeaking* was not explicitly provided in either learning episode, it can be self-derived based on the integrated memory representation. The process is thus an interaction of episodic and

* Corresponding author. Department of Psychology, 36 Eagle Row, Emory University, Atlanta, Georgia, 30322, USA.
E-mail address: pjbauer@emory.edu (P.J. Bauer).
[1] joint first-authorship.

https://doi.org/10.1016/j.learninstruc.2019.101271
Received 12 May 2019; Received in revised form 3 November 2019; Accepted 8 November 2019
0959-4752/

semantic memory (e.g., Tulving, 1972, 1983): information is explicitly taught in separate episodes of learning, and new semantic content that is not tied to a particular episode is derived (see Bauer, Dugan, Varga, & Riggins, 2019, for discussion). Critically, formation of an integrated memory representation is necessary for self-derivation. If participants are taught one member of a pair of related facts but not the other, they do not self-derive new knowledge. The differences in performance are large when both members of the pair of related facts are provided versus only one member of the pair (Cohen's *d* values 0.928, 1.007, and 1.787, for 4-, 6-, and 8-year-old children, respectively: Bauer & Larkina, 2017; see also Bauer & San Souci, 2010).

Among the broader class of inferential processes, self-derivation through memory integration is of special significance because it has a number of features that make it a particularly viable candidate model for accumulation of semantic knowledge. Although other inferential processes share some of these features and thus could in principle inform accumulation of knowledge across learning episodes, their utility on the issue is limited (a) by the stimuli used (e.g., arbitrary associations never meant to be retained: e.g., Zeithamova, Dominick, & Preston, 2012), or (b) because the newly generated information is evanescent (e.g., anaphoric reference: inferences survive in working memory only long enough to ensure comprehension; McKoon & Ratcliff, 1992). In contrast, self-derivation through memory integration is tested using true, factual knowledge, retention of which would be beneficial to achievement. Consistent with this characterization, one important feature of self-derivation through memory integration is that facts derived from integrated episodes are rapidly incorporated into the knowledge base, as revealed by event-related potentials (ERPs; Bauer & Jackson, 2015). Based on a single 400 ms presentation, adults' ERP responses to facts that could be derived from integrated episodes (i.e., "integration facts") were intermediate between ERP responses to facts that were entirely novel and facts that were well known. Based on a second 400 ms presentation, ERP responses to integration facts were no longer distinguishable from responses to well-known facts; ERP responses to both integration and well-known facts were different from those to novel facts (partial eta-squared for different components of the ERP waveforms ranged from .188 to .358; Bauer & Jackson, 2015). Thus facts derived through integration rapidly transition to being treated as "well known."

Another feature of self-derivation through memory integration that makes it an especially viable candidate model for accumulation of semantic knowledge is that the products of self derivation—new facts—are retained over time, at least when tested in the laboratory. Children as young as 4 and 6 years as well as adults remember self-derived facts for at least one week (Varga & Bauer, 2013, 2017b; Varga et al., 2016). Levels of retention are high, with little to no forgetting over the delay. For example, 6-year-olds self-derived novel integration facts on 63% of trials. One week later, they recalled 60% of the facts; the mean levels of performance did not differ from one another (eta-squared 0.007: Varga & Bauer, 2013).

Self-derivation through memory integration as a model for accumulation of semantic knowledge also is supported by the observation that it occurs in elementary school classrooms. In Esposito and Bauer (2017), students in Grades K-3 (ages 5–10 years) were presented with separate yet related episodes of new learning within their classrooms (see also Esposito & Bauer, 2019). The episodes were story passages, each 80–90 words in length. In each story, a character (e.g., Ladybug) engaged in an activity during which it learned a true but novel fact. Later in the same classroom session, students were tested for self-derivation of new facts based on memory integration. Although in the controlled laboratory environment, children as young as 4 years show evidence of self-derivation (e.g., Bauer & Larkina, 2017; Bauer & San Souci, 2010; Varga et al., 2016), kindergarten students exhibited floor levels of performance. In contrast, students in Grades 1–3 performed well, though nominally lower than their counterparts tested in the laboratory (e.g., Bauer et al., 2016a; Bauer & Larkina, 2017).

Further support for self-derivation through memory integration as a viable candidate model for accumulation of semantic knowledge comes from findings that performance on tests of the process relates to academic achievement in both adults and children. For adults, self-derivation through integration relates to both verbal SAT scores and college GPA (*r*s = 0.40 and 0.27, respectively); measures of retention of newly self-derived knowledge over one week predicted academic GPA 2 years later (*r* = 0.34; Varga, Esposito, & Bauer, 2019). For students in Grades 1–3, self-derivation relates to final assessments of reading comprehension (nationally standardized test; *r*s = 0.61) and math performance (classroom grades; *r*s = 0.44). In Grade 3, self-derivation predicts academic performance in math as measured by nationally standardized end-of-course tests (β = 0.54; such tests are not available prior to Grade 3; Esposito & Bauer, 2017).

These findings present a strong case that self-derivation of new factual knowledge through memory integration is a potentially important mechanism by which children may accumulate knowledge over time. However, there are three important unknowns about the process that have implications for its suitability as a model for accumulation of semantic knowledge. First, to date, there have been no tests of whether facts newly self-derived in the classroom are retained over time. Although robust retention is apparent in the laboratory, it cannot be assumed that it also would be observed in the classroom, based on differences in the conditions of encoding and tests of retrieval. In the laboratory, children experience a discrete episode of learning in a unique, distinctive context, which itself may serve as a retrieval cue at the retention test (e.g., Bauer, Stewart, White & Larkina, 2016b). In contrast, in the classroom, children experience multiple hours of instruction in a row, increasing the likelihood of interference. As well, the environment of the retention test does not differentially cue the target learning episode relative to all other episodes of learning in the same context (i.e., cue-distinctiveness is lacking: e.g., Bjork & Richardson-Klavehn, 1989; Herz, 1997; Jacoby & Craik, 1979; Smith, 1988). As a consequence, there is reason to question whether factual information newly self-derived in the classroom is sufficiently robust to be retained and sufficiently accessible to be retrieved. This is a particular concern considering that performance is nominally lower in the classroom than in the laboratory. Accordingly, the first purpose of the present research was to test retention over a 1-week delay of facts self-derived in the classroom through memory integration (Experiment 1).

The second missing element that bears on the suitability of self-derivation through memory integration as a model for accumulation of semantic knowledge is whether the process extends to academic content. Thus far, the materials used in the classroom have been stimuli developed for laboratory testing. They are true facts selected to be previously unknown to and engaging for children (e.g., the Snickers® candy bar is named after a horse). In the present research, across Experiments 1 and 2, we compared self-derivation of new factual knowledge through integration using typical laboratory stimuli (Experiment 1) to self-derivation when the stimuli were facts aligned with the elementary science curriculum (Experiment 2). We used laboratory-based stimuli in Experiment 1 because they are the stimuli for which retention has been demonstrated in the laboratory (e.g., Varga & Bauer, 2013). In Experiment 2, we used the science curriculum both because of its importance to overall educational success (Newcombe, Ambady, Eccles, Gomez, Klahr, Linn, et al., 2009), and because science instruction tends to be cumulative and thus, theoretically, should be especially dependent on integration across related learning episodes. Given that laboratory studies have used a range of stimuli, and no stimulus-set differences have been found, we expected comparable performance across the two types of stimuli.

The third missing element that bears on the suitability of self-derivation through memory integration as a model of accumulation of semantic knowledge is whether children remember details of the episodes in which novel, to-be-integrated facts are conveyed. Clearly, students remember information taught to them in the classroom. They even incorporate misinformation, such as incorrect facts, into their recall (e.g., Butler, Dennis, & Marsh, 2012). In both the laboratory and classroom, children have been found to remember the facts that convey the specific information that, when integrated, serves as the foundation for self-derivation (i.e., stem facts; e.g., *dolphins live in groups called* pods; *dolphins talk by clicking and squeaking*). Stem-fact recall is related to self-derivation such that children who have high levels of self-derivation also tend to have high levels of recall of related stem facts (though the converse is not true; e.g., Bauer & San Souci, 2010). To date there have been no tests of whether children also retain other information from the episodes in which stem facts are embedded, such as themes, characters, and activities. Nor has there been investigation of whether memory for details of explicit learning episodes is impacted by the cognitive operation of integration. Memory for episode details was tested in Experiment 2.

Whether children accurately retain details of explicit learning episodes is an important question in its own right. We typically think of learning episodes as the source of new semantic or world knowledge. Yet they also can serve an important function in episodic and, particularly, autobiographical memory (e.g., Bluck, Alea, Habermas, & Rubin, 2005). That is, individuals remember specific learning episodes as sources of life lessons that they use to direct ongoing and future activity (e.g., Pillemer, 2003; Waters, Bauer, & Fivush, 2014).

The question of whether children accurately retain details of explicit learning episodes also bears on whether the process of cross-episode integration extends beyond the stem facts, to the balance of the information surrounding them (i.e., episode themes, characters, activities). This issue is of theoretical significance for at least three reasons. First, it could be expected to impact memory for the source of information. If integration extends beyond the individual facts to the larger episodes in which they are embedded, then it would be difficult to accurately remember the sources of explicitly taught information (see Butler et al., 2012, for a related argument). The second, related reason is that formation of integrated episodes, with attendant loss of contextual (source) information, would lead to rapid transition of specific episodes to semantic memories (as observed in Bauer & Jackson, 2015). This transition typically is assumed to occur as a result of generalization over multiple episodes (e.g., Rogers & McClelland, 2004). The implication is that, in the case of integrated episodes, the process of "episodic" memories transitioning to "semantic" memories could occur on the basis of a single experience.

The third reason it is important to determine whether the process of integration extends to the entire learning episode is because it bears on *when* integration occurs. Based on studies using ERP and fMRI, it appears that adults integrate memory representations at the time of encoding. In Varga and Bauer (2017a), for example, distinctive ERP patterns were observed at encoding on trials on which adults subsequently self-derived new factual knowledge through integration versus trials on which they failed to self-derive (partial eta-squared = .83; see e.g., Zeithamova et al., 2012, for consistent evidence based on fMRI). In contrast, in children, the processes supporting self-derivation through memory integration seemingly occur on demand, at the time of test. This suggestion is based on findings that when a delay is imposed between encoding of stem facts and test for self-derivation, performance falls substantially. Delay is not the culprit: if the delay is imposed *after* the test for self-derivation, new knowledge is well retained (Varga & Bauer, 2013). This pattern implies that absent the prompt or demand to integrate and self-derive, children did not engage in the process (see also Bauer, King, Larkina, Varga, & White, 2012; and Bauer, Varga, King, Nolen, & White, 2015, for consistent evidence). If this is an accurate depiction, then we would not expect to find evidence of integration unless there is a demand for it. In the self-derivation paradigm, there is a demand to integrate the stem facts (i.e., it is prompted by a question, such as, *How does a pod talk*?), but there is no demand to integrate the larger episode in which the stem facts were embedded. Thus if there is evidence of integration of non-stem-fact content, the process most likely occurred at the time of encoding.

In Experiment 2, we tested students for self-derivation of new science knowledge through integration in the classroom. One week later, we tested whether they remembered specific details of the explicit episodes of instruction, by comparing their recognition of statements that were (a) presented in one or the other of a pair of related episodes ("old"), (b) not presented as part of any episode ("new"), and (c) a "hybrid" of information presented across a pair of related episodes. Endorsement of previously presented statements as "old," and rejection of new statements, would indicate memory for the information surrounding the stem facts. We reasoned that endorsement of hybrid statements as "old" would suggest that the children had integrated the episodes themselves, not only the stem facts featured therein. To address the possibility that children might endorse hybrid statements out of confusion (since they were technically "new" yet comprised of "old" information), we also evaluated children's confidence in their endorsements. If hybrid statements induced confusion, children should be less confident in endorsing them, relative to endorsing old statements and rejecting new statements.

The sites for the present research were second-grade classrooms in a small, rural town in the Southeastern United States. We selected second grade, with children 7–9 years of age, because in prior research, children of this age have performed well on the task of self-derivation through integration (e.g., Esposito & Bauer, 2017, 2019). This makes it likely we would avoid floor performance when testing retention of self-derived information over a 1-week delay. Children of this age also have relatively high levels of memory for the source of information to which they have been exposed (e.g., Cycowicz, Friedman, Snodgrass, & Duff, 2001; Riggins, 2014). This is important in order that endorsement of hybrid statements in Experiment 2 would most likely be the result of cross-episode integration, as opposed to difficulty making source attributions.

In summary, in two experiments, we examined 7- to 9-year-old children's retention of information experienced in the context of tests for self-derivation through integration. In Experiment 1, we tested retention after one week of facts self-derived through integration of separate yet related episodes of new learning. Based on laboratory research with 6-year-olds (Varga & Bauer, 2013), we expected the children would retain newly self-derived facts over the delay. In Experiment 2, we tested retention after one week of other, non-stem-fact information from the episodes in which stem facts were conveyed. We expected that children would accurately recognize information that had been presented in the episodes and correctly reject information that had not been presented. If children integrated the separate episodes, they should accept statements that were a "hybrid" of the separate yet related episodes of new learning. Finally, across experiments, we compared self-derivation of content that was not (Experiment 1) and was (Experiment 2) aligned with the second-grade science curriculum in the school system in which the work took place. We expected to observe self-derivation for both types of stimuli, thus demonstrating the generalizability of findings from the laboratory to classroom curricula.

2. Experiment 1

2.1. Method

2.1.1. Participants

The participants were 96 (56 female) students in second grade classrooms in the same school in a public school system (M = 8.11 years; range = 89–109 months). Consent forms were sent home through parent folders (the typical means of communication between the school system and students' parents/guardians). Only the data from children whose parents/guardians returned signed consent forms were included in analyses (approximately 39% of the population). The sample was thus the population of children for whom parents/guardians had provided consent for use of their data. Based on the results of prior research in which effect sizes of 0.3 (and greater) have been observed (e.g., Bauer & Larkina, 2017; Varga & Bauer, 2013), a sample of 45 would provide adequate power (0.8) for the planned repeated measures design. Thus the study is sufficiently powered.

Reflecting the diversity of the community, based on parental report, the sample was 38% African-American, 29% Caucasian, 22% Hispanic/Latinx, 9% multiracial, and 2% unreported. Eighty-four percent of children in the community qualify for federally funded school lunch assistance. Of the 82 participants whose families reported caregiver education, 35% had a high school education or less, 28% had some training beyond high school, 15% had a technical or associates degree, 16% had a college bachelor degree, and 6% had education beyond a college degree. Participating teachers were thanked with a $20 gift card, parents were thanked with a $10 gift card, and participating children were thanked with a small school supply item (e.g., eraser). The Institutional Review Board and the School Board of the participating school system reviewed and approved all study protocol and procedures for this and the second experiment.

2.1.2. Stimuli

The stimuli were eight novel "stem" facts, each of which was a true fact. The eight stem facts formed four pairs such that, within a pair, the two facts were related and could be combined to generate a novel integration fact. One pair of stem facts was that (a) *tigers are the largest cats*, and (b) *the largest cats swim to cool off*. These facts could be combined to support self-derivation of the new knowledge that *tigers swim to cool off*. The stimuli were pilot tested to ensure both that the stem and integration facts were novel to children in the target age range and that both stem facts were necessary for production of the integration facts. Pilot testing was conducted in the laboratory in small groups to mimic the conditions of testing in the classroom.

The stem facts were featured in text passages resembling picture stories. The passages were 81–89 words in length, distributed over 4 pages. Each page featured a hand-drawn illustration depicting the main actions of the text; the text was not included on the page. For example, the story of the "Contest of the Cats" was rendered as (the stem fact is indicated in italics):

> Page 1. Frog knew that there are many large cats in the world. One day, she held a contest to find out which cat was the biggest.
> Page 2. Frog was the judge. Each type of cat lined up so she could see their sizes.
> Page 3. There were many big cats, but *the tiger was the largest cat in the world.* Frog happily announced the winner to everyone.
> Page 4. The contest was complete and now Frog knew that *tigers are the largest cats in the world.*

All of the passages were similar in structure: a character learned a true but novel fact in the course of a short story. Each passage had a different animal as the main character. Each pair of passages was drawn from a different domain: cats, the Queen of England, apricots, and Snickers®. The stem facts were presented on Page 2 or 3 of the passages and were repeated on the final page; the integration facts were not presented. Following Esposito and Bauer (2017), the text passages were presented in digital book format. Each illustration was scanned into a PowerPoint® slide. The audio portions were recorded by a native English speaker.

2.1.3. Procedure

Children participated in two sessions in their schools, approximately one week apart (*M* delay = 6.19 days, *SD* = 0.80). In Session 1, they were tested for self-derivation of new factual knowledge through integration of separate episodes. The protocol was administered in the children's classrooms to the entire class (approximately 20–23 children per classroom). Only the data from children whose parents/guardians returned signed consent forms were included in analyses. One week later, children were tested for memory for the integration facts. Testing was conducted one-on-one by one of seven undergraduate research assistants. All assistants had been extensively trained to administer the protocol in the same manner. Fidelity of administration was assured by one of the authors, who monitored the assistants throughout protocol administration. There were no protocol errors on the measure of interest in the present research.

2.1.3.1. Session 1. The 45-min classroom sessions were divided into three phases. In Phase 1, children heard the first member of each of the four pairs of text passages (i.e., one passage each about cats, the Queen of England, apricots, and Snickers®). Children then engaged in an unrelated buffer activity for approximately 10 min. Phase 2 commenced after the buffer activity. The children heard the second member of each stem-fact passage pair (i.e., the second passage each about cats, the Queen of England, apricots, and Snickers®). Following Phase 2, children engaged in a second 10 min buffer activity. For both phases, the illustrations conveying the main actions of the passages were projected onto the classroom screen (approximately 4′ by 6'). The pre-recorded audio tracks were played through speakers. The slides and audio were advanced automatically, ensuring consistent timing across classrooms. The text passages within domains were counterbalanced and domains were presented in one of four pre-determined random orders; each order was used approximately equally often across classrooms.

In Phase 3, children were tested for self-derivation of new factual knowledge through integration of the members of the pairs of related stem facts, first in open-ended and then in forced-choice format. Both open-ended and forced-choice formats were used to guard against potential floor effects in open-ended testing and thus ensure adequate variability for analyses. Children also were tested for recall (open-ended) and recognition (forced-choice) of the stem facts. For example, to test self-derivation of integration facts in open-ended format, children were posed the question "What do tigers do to cool off?" (which could be answered through integration of the stem fact from the sample passage above [*tigers are the largest cats in the world*] with the stem fact in the paired passage, not presented [*the largest cats in the world swim to cool off*]). In forced-choice format, they were presented the same question along with three choice alternatives, one of which was correct: (a) swim, (b) shower, (c) sleep. To test open-ended recall of stem facts, children were posed the question "What is the largest cat in the world?" In forced-choice format, they were presented the same question along with three choice alternatives, one of which was correct: (a) jaguars, (b) tigers, (c) cheetahs.

All questions were read aloud by a researcher. For open-ended testing, children recorded their responses on an answer sheet. They first were tested for self-derivation of the integration facts, followed by recall of the stem facts. After open-ended testing, children were given response devices (i.e., "clickers") to record their answers to 3-alternative forced-choice questions (one correct alternative and two distracters; chance = 33%). As in open-ended testing, the integration questions were posed first, followed by the stem fact questions. The integration and stem-fact questions were presented in one of four pre-determined random orders; each order was used approximately equally often across classrooms and text passage orders.

2.1.3.2. Session 2. Approximately one week after the self-derivation test, children were tested for recall followed by recognition of the integration facts they were expected to self-derive at Session 1. Testing took place in a different classroom. Because the setting, format, and test administrators were different from Session 1, children first were asked two questions about the stories presented one week previously to remind them of the material in which we were interested; the integration facts were not prompted. After the story-reminder questions, children were asked open-ended questions testing recall of the integration facts (e.g., "How does a pod talk?"), followed by forced-choice testing if they failed to produce a correct open-ended response (3 alternatives; chance = 33%). The forced-choice alternatives were the same as those used at Session 1.

2.1.4. Scoring

At Session 1, children received 1 point for each correct response to a self-derivation question, for a total possible score of 4 in each of open-ended and forced-choice testing of the integration facts. They could score up to 8 in each of open-ended and forced-choice testing of the stem facts (i.e., 2 stem facts for each integration fact). At Session 2, children received 1 point for each integration fact recalled and 1 point for each additional integration fact recognized in forced-choice. Thus children could score up to 4 points for recall in open-ended testing and 4 points for their total score (open-ended recall plus additional unique items in recognition).

2.2. Results

The results are presented in three sections: (a) self-derivation of integration facts and stem-fact memory performance at Session 1, (b) recall and recognition of the integration facts one week later at Session 2, and (c) relations between self-derivation (and stem-fact memory) and later memory for the integration facts.

2.2.1. Self-derivation and stem-fact memory at session 1

The distribution of scores at Session 1 is depicted in Fig. 1. In open-ended testing, children self-derived the integration facts on a mean of 1.53 trials (*SD* = 1.06; max = 4) and they recalled the stem facts on a mean of 4.56 trials (*SD* = 2.00; max = 8). On the balance of the trials, children indicated that they “didn't know” or left the answer sheet blank; they rarely provided a content response that was incorrect. In forced-choice testing of integration facts, children selected the correct answers from among distracters on a mean of 3.27 trials (*SD* = 0.95; max = 4) and they selected the correct answers to the stem-fact questions on a mean of 6.78 trials (*SD* = 1.67; max = 8). For both integration and stem facts (0.82 and 0.85, respectively), forced-choice accuracy was significantly above chance (0.33) for a mean difference > 0.49, 95% CIs[0.44 to 0.56], $t(95)$s > 20.21, $ps < .001$. Thus the children were relatively successful at the task.

2.2.2. Recall and recognition of integration facts at session 2

One week after the test for self-derivation of integration facts, children had high levels of recall and recognition of them, as depicted in Fig. 2. In open-ended testing, children recalled the self-derived

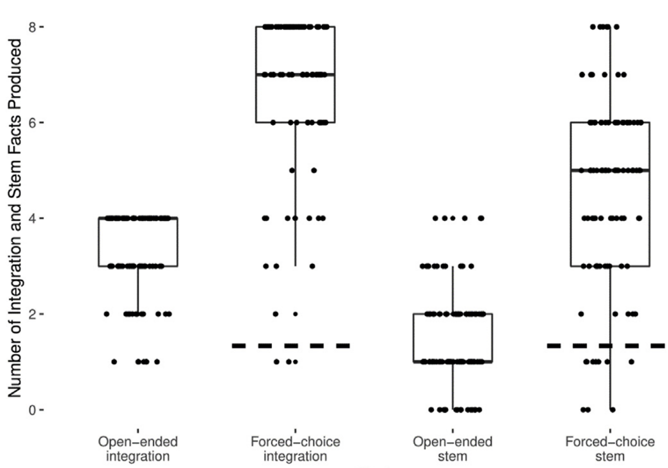

Fig. 1. Number of integration and stem facts produced in open-ended and forced-choice testing in Experiment 1. The dashed lines in forced-choice testing indicate the level of performance expected by chance.

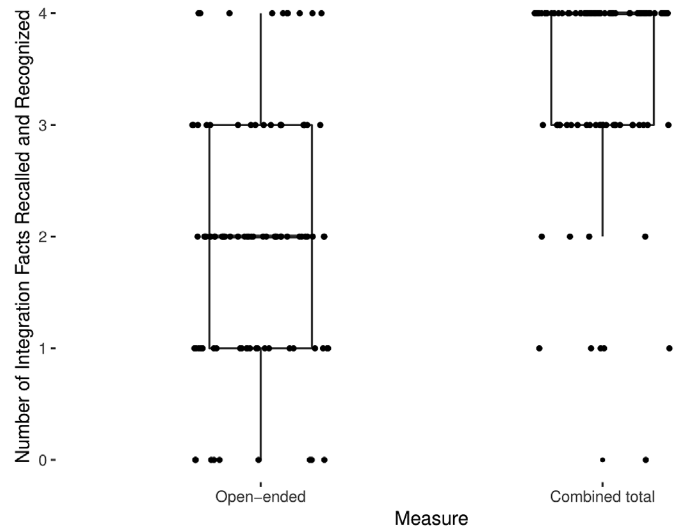

Fig. 2. Number of integration facts recalled (open-ended) and recalled and recognized (combined total) at Session 2 in Experiment 1.

integration facts on a mean of 1.95 (*SD* = 1.12) trials (max = 4). This represents a statistically significant *increase* in performance from Session 1 by a mean of 0.42, 95% CI[0.21 to 0.62], $t(95) = 4.03$, $p < .001$, $d = 0.40$. For trials on which the children did not recall the integration facts, they were given the opportunity to choose the correct answer from among distractors. Children recognized a mean of an additional 1.57 (*SD* = 0.86) integration facts, for a total combined recall/recognition score of 3.45 (*SD* = 0.88; max = 4). Thus students successfully retained the information over the 1-week delay. Indeed, consistent with findings from laboratory-based research (Varga & Bauer, 2013), there was no loss of information over the delay.

2.2.3. Relations between stem- and integration-fact performance and later memory

Table 1 summarizes the correlations between self-derivation performance in open-ended and forced-choice testing formats and recall and recognition of the stem facts (open-ended, forced-choice, respectively) at Session 1, and their predictive relations with recall and recognition of the integration facts at Session 2. There was a consistent pattern of positive relation among the variables. Two aspects of the interrelations are especially noteworthy. First, as observed in prior research (Bauer & San Souci, 2010), within Session 1, open-ended self-derivation was related to recall of the stem facts ($r = 0.66$). The same relation was observed when both self-derivation and stem fact memory were tested using forced-choice ($r = 0.58$). These relations are depicted in Appendix A, Figure A.1, Panel A. Second, performance at Session 1—both self-derivation and stem-fact memory—was predictive of recall and recognition of self-derived knowledge at Session 2. The correlations among the measures ranged from 0.44 to 0.57. Appendix A, Figure A.2, Panel A provides depictions of the most prominent of these relations (between open-ended self-derivation of integration facts and recall of stem facts at Session 1 and open-ended recall and total memory for integration facts at Session 2). R^2 values indicate that self-derivation and stem-fact memory at Session 1 accounted for between 19% and 32% of the variance in retention of self-derived knowledge over the 1-week delay.

2.3. Discussion

In a replication of Esposito and Bauer (2017), second-grade students self-derived new factual knowledge through integration of separate yet related episodes of new learning in their classrooms. The present research also extended prior research by testing retention from classroom

Table 1
Pearson correlations among measures of self-derivation performance and stem-fact memory in experiment 1.3

		Measure					
		1	2	3	4	5	6
1: Open-ended self-derivation Session 1	*r*	1	.66**	.43**	.46**	.56**	.47**
	(N)	(96)	(95)	(96)	(96)	(96)	(92)
2. Open-ended stem fact Session 1	*r*		1	.53**	.55**	.57**	.55**
	(N)		(95)	(95)	(95)	(95)	(91)
3. Forced-choice self-derivation Session 1	*r*			1	.58**	.47**	.44**
	(N)			(96)	(96)	(96)	(92)
4. Forced-choice stem Session 1	*r*				1	.47**	.50**
	(N)				(96)	(96)	(92)
5. Recall of integration facts Session 2	*r*					1	.62**
	(N)					(96)	(92)
6. Total memory integration facts Session 2	*r*						1
	(N)						(92)

learning through integration. The results were clear: children had high levels of recall and recognition of their self-derived knowledge one week later. Indeed, performance after the 1-week delay was statistically higher than performance at Session 1. The findings thus demonstrate that productive extension of knowledge through memory integration is a viable model for accumulation of knowledge not only in the laboratory (e.g., Bauer & Larkina, 2017), but in the classroom as well.

The results of the present study make clear that children remember information they themselves self-derived, through integration of separate yet related episodes of new learning. Successful self-derivation depends on formation of a memory representation that is the integration of separate story episodes (Bauer & Varga, 2017). This raises the question of the consequences of integration for memory of the episodes themselves. It is possible that integration extends to the episode as a whole, creating a representation that is a blend or hybrid of the individual episodes on which it is based. In this case, children may fail to distinguish statements that were actually presented in the stories from statements that were never presented, but which incorporate elements from both stories. Alternatively, it is possible that integration is restricted to the stem facts themselves and does not extend to the episode as a whole. In this case, we would expect preservation of distinct episode boundaries, permitting children to distinguish statements that were actually presented in the stories from statements that were never presented, even if those new statements are "hybrids" created from the pair of episodes. We tested these competing possibilities in Experiment 2. We used as stimuli facts derived from the children's science curriculum, thus permitting a between-experiment comparison of self-derivation through integration when stimuli were based on those used in the laboratory (Experiment 1) with self-derivation when the stimuli were aligned with the elementary science curriculum (Experiment 2).

2.4. Experiment 2

Experiment 2 had two major purposes. First, to evaluate self-derivation through memory integration as a model for classroom learning, we used stimuli that were generated from the state standards for second-grade science curriculum in the host school system. The stimuli were generated from material that had not yet been covered in the classroom. All other conditions of testing for self-derivation through integration were the same across experiments.

The second major purpose of Experiment 2 was to test whether children remember details of the explicit learning episodes in which to-be-integrated stem facts are conveyed. This question stands to inform whether cross-episode integration extends beyond the stem facts to the episodic information surrounding them. To ensure a robust test of the question, in a between-subjects manipulation, we used two different levels of surface-feature similarity: (a) high surface-similarity, in which the two passages in a pair had the same main character (e.g., a lizard in both passages); and (b) low surface-similarity, in which the passages within a pair had different main characters (e.g., a lizard in one passage and a cat in the other). We reasoned that the same character across the two passages might make it more likely that integration would extend beyond the stem facts to the entire story episodes, whereas different characters might make it more likely that the separate episodes would remain distinct.

2.5. Method

2.5.1. Participants

The participants were 103 (46 female) children in second grade (M = 8.17 years; range = 91–112 months). The children were drawn from the same population as Experiment 1. Testing was conducted two years after Experiment 1, and none of the children had taken part in Experiment 1. The consent procedure was the same as in Experiment 1. Only the data from children whose parents/guardians returned signed consent forms were included in analyses (approximately 41% of the population). As was the case in Experiment 1, the result was a population sample. As discussed in Experiment 1, the study is sufficiently powered: a sample of 45 would provide adequate power (0.8) for the planned repeated measures design.

Reflecting the diversity of the community, based on parental report, the sample was 37% African-American, 31% Caucasian, 20% Hispanic/Latinx, 5% multiracial, 1% American Indian/Alaska Native, and 6% unreported. Of the 95 participants whose families reported caregiver education, 46% had a high school education or less, 22% had some training beyond high school, 13% had a technical or associates degree, and 19% had a college bachelor degree. Children participated in two sessions approximately one week apart (M delay = 7.23 Days, SD = 0.96). Participating parents, teachers, and children were thanked as in Experiment 1.

2.5.2. Stimuli

As in Experiment 1, the stimuli were novel facts. Related pairs of facts could be integrated with one another and serve as the basis for self-derivation of a novel integration fact. The stem facts were derived from domains covered in the children's science curriculum: matter, sound, space, and life cycles. The specific facts had yet to be covered in the classroom. Prior to their administration in the classroom, the facts were pilot tested to ensure both they were unknown to children in the target age range, and that exposure to both members of the pairs of related stem facts was necessary for generation of the integration facts.

As in Experiment 1, the stem facts were featured in illustrated text passages 81 to 89 words in length, distributed over 4 pages; the text was not featured on the page. To test whether the surface similarity of the passages impacted preservation of distinct episodic features, for approximately half of the children (n = 43), the two passages in a pair had the same animal as the main character, and for the other half of the children (n = 60), the two passages in a pair had different animals as the main characters.

As in Experiment 1, only the stem facts were included in the passages; the integration facts were not presented. Also as in Experiment 1, the text passages were presented in digital book format. Each illustration was scanned into a PowerPoint® slide. The audio portions were recorded by a native English speaker.

We also developed 24 memory stimuli to be used at Session 2, 8 of each of three types (see 3.2.3.). An illustration of each of the three different types of stimuli is provided in Fig. 3. One third of the statements (n = 8) were taken directly (verbatim) from one of the story passages. There was one verbatim statement from each passage (thus half of the statements were from each passage in a pair of related passages). One third of the statements were entirely new—they featured the same characters, themes, and settings as the stimulus story passages, but they had not been presented verbatim and neither were they "gist" representations of statements in the stories. There were two new statements for each story passage pair. The remaining one third of statements were hybrids created by combining elements of the two story passages in a pair. That is, a portion of the statement came from Passage 1 (in the example in Fig. 5, Panel A: *rocket*) and a portion of the statement came from Passage 2 (*for Lizard's birthday, her dad gave her*). The two portions were combined to create a hybrid statement (e.g., Panel B: *Lizard's dad got her a rocket for her birthday*). Finally, we developed a Likert-type scale to assess children's confidence in their endorsements of the statements as "old" (see 3.2.3.). The scale was 4 points, with 1 indicating "not sure at all," 2 indicating "a little sure," 3 indicating "mostly sure," and 4 indicating "completely sure."

2.5.3. Procedure

Children were tested for self-derivation of new factual knowledge through integration of separate episodes in Session 1. One week later, they were tested for recognition of three different types of statements.

2.5.3.1. Session 1. The Session 1 protocol was the same as in Experiment 1. Although the protocol was administered to the entire class, only the data from children whose parents/guardians returned signed consent forms were included in analyses.

2.5.3.2. Session 2. Approximately one week after presentation of the stem-fact stories and the test for self-derivation, children were tested for recognition of 24 statements related to the story passages. One third of the statements (n = 8) were taken verbatim from one of the story passages; half of the statements were from each passage in a pair of related passages. To the extent that the children remembered the story passages, they were expected to indicate that they had heard these statements before (i.e., to indicate that they were "old"). One third of the statements (n = 8) were entirely new. To the extent that the children remembered the story passages, they were expected to indicate that they had not heard these statements before (i.e., to indicate they were "new"). The remaining one third of statements (n = 8) were hybrids created by combining elements of the two story passages in a pair (see Fig. 3). If children maintained the boundaries of the episodes even after integrating the story passage information, they were expected to indicate they had not heard these statements before (i.e., to indicate they were "new"). However, if integration extended beyond the stem facts to the entire episodes, resulting in a blended representation, then because these statements featured content from both stories, children were expected to indicate that they had heard these statements before (i.e., to indicate they were "old").

Children were told that "Last week, you heard some stories in your classroom. I am going to read you some sentences and I want you to tell me if you heard this information in the stories last week." Testing was forced-choice, with two alternatives: the child endorsed having heard the statement in the context of the story passage paradigm one week earlier or the child had not heard the statement. After each trial, children were asked to rate "how sure" they were of their answer using a 4-point Likert-type scale. Testing was conducted one-on-one by one of 12 research assistants (including one of the authors). All assistants had been extensively trained to administer the protocol in the same manner. Fidelity of administration was assured by another of the authors, who monitored the assistants throughout protocol administration. There were no protocol errors on the measure of interest in the present research.

2.5.4. Scoring

At Session 1, children received 1 point for each correct response. Thus they could score up to a 4 in each of open-ended and forced-choice testing of the integration facts. They could score up to 8 in each of open-ended and forced-choice testing of the stem facts. At Session 2, children received one point for each statement endorsed as "old" (max = 8 for each statement type: old, new, hybrid). For the confidence scale, children received a score of 0 for "not sure at all," 1 point for "a little sure," 2 points for "mostly sure," and 3 points for "completely sure."

2.6. Results

The results are presented in three sections: (a) self-derivation of the integration facts at Session 1 and memory for the stem facts, (b) endorsement of statements as "old" at Session 2 and confidence in the endorsements, and (c) relations between self-derivation at Session 1 and patterns of and confidence in endorsement of statements as "old" at Session 2.

2.6.1. Self-derivation and stem-fact memory at session 1

To determine whether self-derivation through memory integrations extends to classroom science content, we examined levels of self-

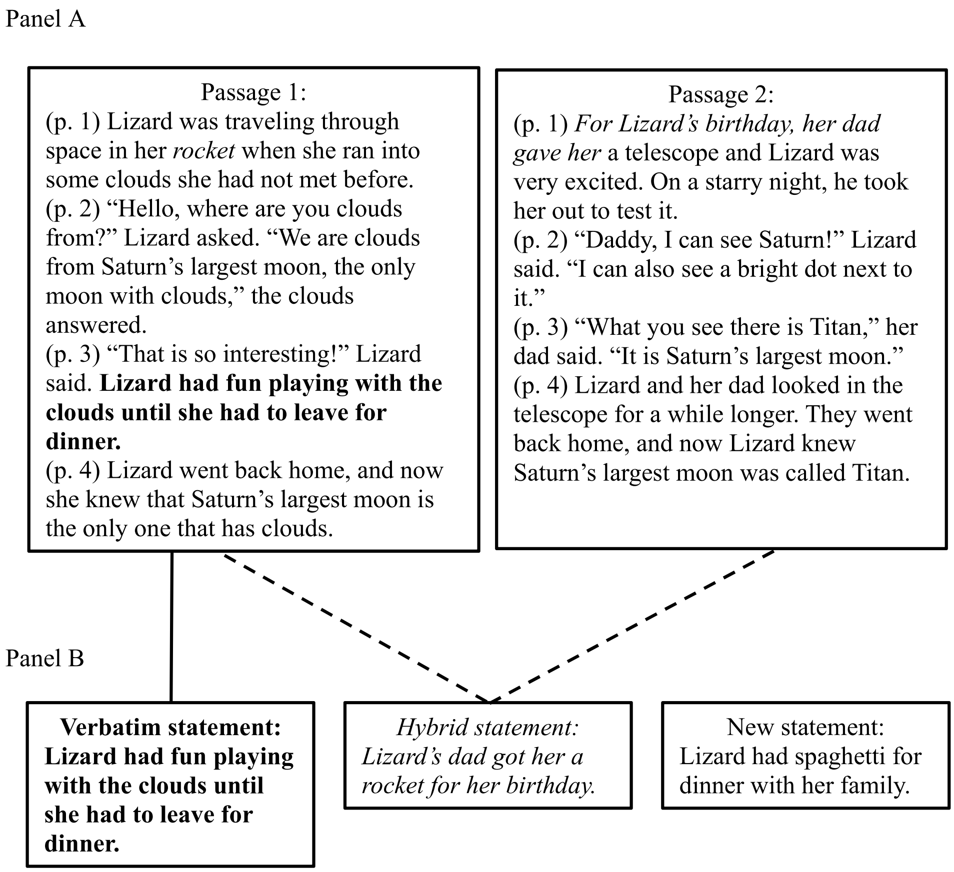

Fig. 3. A pair of related episodes (Passage 1 and Passage 2) presented at Session 1 (Panel A) and illustrations of each of three statement types tested at Session 2 (Panel B), used in Experiment 2. Information presented in **bold** in Passage 1 illustrates the verbatim statement type: the statement was taken directly from one of the passages the children experienced at Session 1. Information presented in *italics* in Passages 1 and 2 was combined to create a hybrid statement: all of the content in hybrid statements had been presented in the passages; a portion of the statement was presented in Passage 1 and a portion was presented in Passage 2. New statements were not represented in the learning episodes, but were thematically consistent with them.

Table 2
Descriptive Statistics for Self-derivation and Stem-fact Performance in Experiment 2.

	Task and Testing Phase			
	Integration Facts		Stem Facts	
	Open-ended	Forced-choice	Open-ended	Forced-choice
Similarity condition	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
High similarity	1.59 (1.29)	2.82 (1.19)	5.00 (1.93)	5.62 (1.62)
Low similarity	1.42 (1.26)	3.09 (0.88)	4.60 (2.12)	5.45 (1.88)
Overall	1.49 (1.25)	3.00 (1.02)	4.72 (2.08)	5.52 (1.84)

Note: *M* refers to mean correct performance and (*SD*) refers to the standard deviation.

derivation in both open-ended and forced-choice formats, as well as recall and recognition of the stem facts. Statistics describing performance in each of the similarity conditions, as well as overall (across similarity conditions), are provided in Table 2. A MANOVA examining differences in self-derivation and stem fact performance revealed no difference in open-ended or forced-choice performance across the similarity conditions, $F(92) = 1.30$, $p = .28$, $\eta^2 = 0.06$. In light of the absence of difference between similarity conditions, we examined forced-choice performance across the similarity conditions. For both integration facts (0.75 correct) and stem facts (0.69 correct), forced-choice performance was significantly above chance (0.33) with a mean difference > 0.36, 95% CI[0.31 to 0.47], $t(99)\text{s} > 15.61$, $ps < .001$.

To determine whether self-derivation through memory integration differed for science and non-science content, we compared performance in the present experiment with that in Experiment 1. For ease of comparison, performance is depicted in Fig. 4 for both experiments. Notably, as suggested by inspection of Panel A, levels of open-ended self-derivation of the integration facts and open-ended recall of the stem facts did not differ between the experiments, even though each used entirely different stimuli, $t(189) = 0.22$, $p = .83$, $d = 0.03$, and $t(188) = -0.53$, $p = .59$, $d = 0.08$ (respectively). In forced-choice testing (Panel B), performance in Experiment 1 was nominally higher than in Experiment 2, though the difference did not reach statistical significance, $t(194) = 1.93$, $p = .06$, $d = 0.28$. The difference in recognition of the stem facts was statistically significant, $t(194) = 5.01$, $p < .001$, $d = 0.72$, with higher performance in Experiment 1. This across-experiment comparison makes clear that children self-derive new knowledge through memory integration across a wide range of stimuli, including materials derived from the science curriculum. Possible reasons for lower levels of forced-choice selection are discussed in Section 3.4.

As observed in Experiment 1, in the present experiment, memory for stem facts and self-derivation through integration were positively correlated in both the open-ended, $r(93) = 0.79$, $p < .001$, and forced-choice, $r(98) = 0.58$, $p < .001$, testing formats. Depictions of these relations are provided in Appendix A, Figure A.1, Panel B.

2.6.2. *Test for recognition and confidence at session 2*

To address the questions of whether (a) children remember information from the larger episodes in which stem facts are presented (patterns of endorsement of “old” and “new” statements), and (b) integration extends beyond the stem facts to the entire episodes in which they are presented (patterns of endorsement of “hybrid” statements), we conducted a 2 (surface similarity: high, low) x 3 (statement type: old, new, hybrid) mixed analysis of variance, with repeated measures on statement type. Contrary to predictions, patterns of endorsement did not differ as a function of whether the members of the stem-fact passage pairs had the same or different main characters ($Fs < 0.05$, $ps > .83$, $\eta^2 < 0.001$). In subsequent analyses, we collapsed across levels of surface similarity.

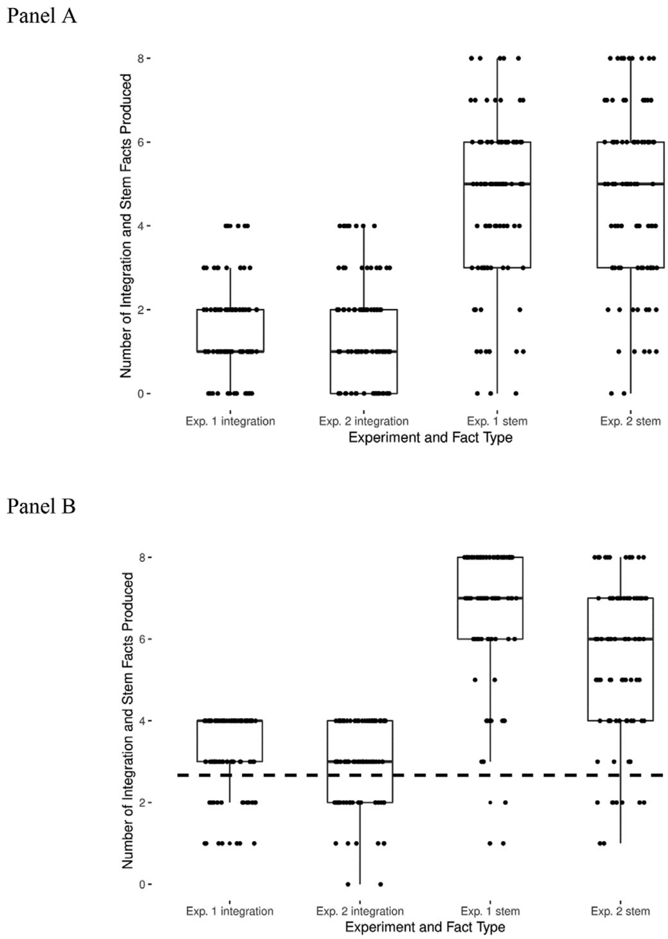

Fig. 4. Number of integration facts produced and stem facts recalled in open-ended testing in Experiments 1 and 2 (Panel A) and number of integration facts selected from among distracters and stem facts recognized in forced-choice testing in Experiments 1 and 2 (Panel B).

As suggested by inspection of Fig. 5, Panel A, there was a main effect of statement type, $F(2, 100) = 92.39$, $p < .001$, $\eta^2 = 0.65$. Follow-up pairwise comparisons with Bonferroni corrections revealed significant differences in children's rates of endorsement of all three statement types. Specifically, children more often indicated they had heard the old statements than both new and hybrid statements; hybrid statements were endorsed at higher levels than new statements.

Analyses against chance performance indicated a systematic pattern of endorsement of old statements, such that children responded that they had heard old statements at a level that exceeded chance (.50), $t(102) = 8.66$, $p < .001$. The mean difference from chance was 0.20 (95% CI[0.16 to 0.25]). In contrast, for both the new and hybrid statement types, children had below-chance levels of endorsement, indicating accurate recognition that these statements had not been presented in the stories, $ts(102) = 7.36$ and 3.61, $ps < .001$ (respectively). The mean differences from chance were -0.17 and -0.10 (95% CIs [-0.21 to -0.12] and [-0.14 to -0.04], respectively.

We also examined the confidence judgements children provided regarding their decisions whether to endorse statements as “old.” If children were confused by the hybrid statements, we expected they would have lower confidence judgements regarding those statements compared to old and new statements. To address this possibility, we conducted a one-way repeated measures analysis of variance (ANOVA) with statement type (old, new, hybrid) as a within-subjects factor. As suggested by inspection of Fig. 5, Panel B, the ANOVA revealed a main effect of statement type, $F(2, 98) = 48.30$, $p < .001$, $\eta^2 = 0.33$. Follow-up comparisons with Bonferroni corrections revealed that children were significantly more confident in their judgments for old statements than new or hybrid statements. Children also were significantly more confident in accepting hybrid statements than new statements. This pattern is not what would be expected if children were simply confused by the hybrid statements (in which case, their confidence scores would have been lowest overall).

2.6.3. *Relations between self-derivation and endorsement of statements as “old” and confidence judgments*

Because self-derivation performance on each of the four trials was nested within each child, we used multi-level modeling (MLM) to test relations between self-derivation and stem-fact performance and endorsement of statements as “old.” Multi-level modeling allows for individual intercepts and controls for the nesting of data within individuals. Specifically, it permits predictors at the level of the individual trial and at the level of the person. It also takes into account the interdependency of multiple observations per person (i.e., 4 trials for each child), correcting for the biases in parameter estimates resulting from dependency of the observations (Raudenbush & Bryk, 2002; Wright, 1998). Because confidence judgments also were nested within each child, we used MLM to examine relations between children's confidence in their judgments and their patterns of endorsement of the statements.

2.6.3.1. Self-derivation and endorsement of statements as “old”. To minimize the complexity of the analysis, we conducted separate multi-level models for each statement type (old, new, hybrid; see Table 3, Panel A, for unchanged null models). We first tested null models to determine whether there was variance at both levels of the model: Level 1 (trials) and Level 2 (individuals; e.g., Nezlek, 2001; Raudenbush & Bryk, 2002). The results revealed sufficient variance at both levels for all three statement types, with 23%, 21%, and 31% of variance between person for old, new, and hybrid statements (respectively). The full analytic approach is described in Appendix B.

We next examined whether self-derivation performance in Session 1 on each trial predicted endorsement for statements related to that specific trial. As reflected in Table 3, Panel B, self-derivation performance accounted for significant variance in endorsement of old statements. That is, on trials on which children integrated the pairs of related stories, as evidenced by successful self-derivation, they were more likely to correctly endorse old statements as “old.” In contrast, self-derivation performance did not account for significant variance in endorsement of new or hybrid statements. Thus on any given trial, whether the children integrated the pairs of related stories (as evidenced by successful self-derivation) was not related to the likelihood that they would accept new or hybrid statements as old.

2.6.3.2. Confidence judgments and endorsement of statements as “old”. We also conducted analyses to determine whether children's confidence in their judgments related to their patterns of endorsement of the statements. As in the previous analyses, to minimize the complexity, we conducted separate multi-level models for each statement type (old, new, hybrid). The null models were unchanged from above (see Table 3, Panel A).

In all three models, confidence judgements were significantly and positively related to endorsement of the statement type (Table 3, Panel C). Thus for all three statement types, children expressed higher confidence when they accepted a statement as coming from the story. This analysis provides additional evidence that children were not simply confused by the hybrid statements—metacognitively, they treated hybrid statements no differently than old and new statements.

2.7. *Discussion*

The findings on self-derivation of new factual knowledge through memory integration in the present experiment replicate those from

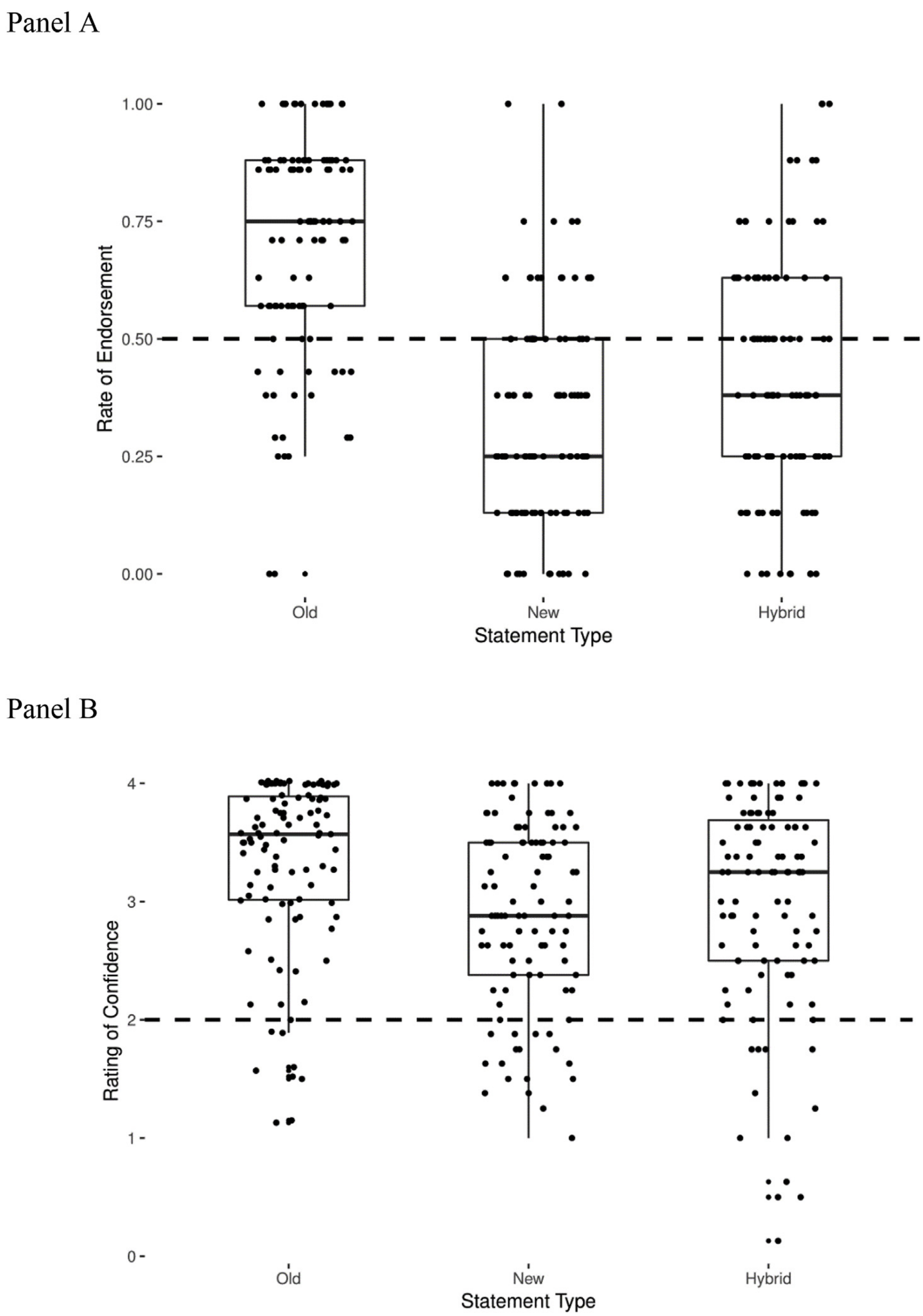

Fig. 5. Endorsement of statements as "Old" as a function of statement type (Panel A) and ratings of confidence in endorsement of statements as "Old" as a function of statement type (Panel B) in Experiment 2.

Experiment 1. The children self-derived new factual knowledge in open-ended testing and selected the correct answers from among forced-choice options. Open-ended self-derivation performance did not differ from that in Experiment 1, even though in the present experiment, the content over which the children operated was derived from their science curriculum. These findings bear on the suitability of self-derivation through memory integration as a model of classroom learning, demonstrating that the process extends to academic content. Children's open-ended stem-fact recall also did not differ between experiments. However, in forced-choice testing, children in Experiment 1 selected correct answers to integration and stem-fact questions more frequently than children in Experiment 2, though the difference for integration facts was not statistically significant. We speculate that this may be due to the nature of the distracters. In Experiment 1, the distracters were familiar terms (e.g., Question: "What animal is the most popular chocolate bar in the world named after?" Answer choices: "a horse, a pig, a sheep.“). In contrast, in Experiment 2, the distracters were less familiar scientific terms (e.g., Question: "What is the name of Saturn's largest moon?" Answer choices: "Predator, Titan, Epperson.“) Nevertheless, in this experiment, as in Experiment 1, forced-choice selection of integration and stem facts was reliably greater than chance.

The results from Session 2 make clear that the children remembered details of the story episodes that were the basis for self-derivation of the integration facts. They correctly indicated that story statements that had been presented were "old," and that story statements that had not been presented in any form were "new." Rates of endorsement of the hybrid statements were intermediate between those for old and new statements; they nevertheless were significantly below chance. Thus

Table 3
Multi-level modeling (MLM): Null model (panel A), and analyses of endorsement of statements (B) and confidence judgments (panel C) in experiment 2.

Panel A: Null model				AIC	BIC	-2LL
	τ_{00}	σ^2				
Statement type						
Old	23%	77%		180.92	192.35	−87.46
New	21%	79%		296.94	308.91	−145.47
Hybrid	31%	69%		296.92	308.83	−145.46
Panel B: Self-derivation as a predictor of statement endorsement						
	γ_{10}	*t*	*p*	AIC	BIC	-2LL
Statement type						
Old	0.11	3.00	.003	156.63	171.55	−74.32
New	0.04	0.95	.34	278.94	294.57	−135.47
Hybrid	−0.02	−0.45	.66	280.92	296.46	−136.46
Panel C: Confidence judgement as a predictor of statement endorsement						
	γ_{10}	*t*	*p*	AIC	BIC	-2LL
Statement type						
Old	0.27	15.33	< .001	6.36	21.6	0.82
New	0.13	6.91	< .001	255.84	271.8	−123.92
Hybrid	0.17	8.82	< .001	229.19	245.05	−110.59

Note: AIC = Akaike's information criterion; BIC = Schwarz's Bayesian criterion; -2LL = -2log-likelihood value, or the deviance of the log-likelihood.

children reliably indicated that they had not heard the information in the hybrid statements the week before. Especially against the backdrop of systematic acceptance of old statements and rejection of new statements, this effect is meaningful. It suggests that children largely preserved the boundaries of the story episodes and thus that they were not integrated with one another. The patterns of findings did not differ for stories with higher versus lower levels of surface similarity.

3. General discussion

The present research had three primary purposes. Experiment 1 was a test of the first purpose, which was to determine whether elementary-age children (8-year-olds) retain factual information they have self-derived in the classroom based on integration of separate yet related episodes of new learning. The result was clear—the children had high levels of recall and recognition of facts they had self-derived one week earlier. Indeed, children's open-ended recall of the integration facts at Session 2 was statistically significantly higher than their open-ended self-derivation of them one week earlier. The increase in performance can be attributed to the research design in which, at Session 1, following open-ended testing, children were given the same test prompts, this time in three-alternative forced-choice format. Correct selection in forced-choice testing could result in retrieval-based learning (see Fazio & Marsh, 2019, for a review), facilitating retention and subsequent recall after the delay. Consistent with this interpretation, forced-choice selection of the integration facts at Session 1 correlated with open-ended recall of the integration facts one week later ($r = 0.47$). The correlation was not statistically significantly different than that between open-ended self-derivation (Session 1) and open-ended recall (Session 2; $r = 0.56$). In summary, Experiment 1 demonstrated retention over of a delay of factual information self-derived in the classroom by elementary-age children.

The second major purpose of the present research was to test whether classroom-based self-derivation through memory integration extends to academic science content. In prior research in elementary classrooms (Esposito & Bauer, 2017), children have been tested on stimuli designed for and used in a laboratory setting. The stimuli were true facts selected for their novelty and likely interest to children in the target age range, and to ensure that the integration facts could not be generated without exposure to both members of the stem-fact pairs. Yet if self-derivation through memory integration is a good model for classroom learning, it must be tested with academically-relevant content. Accordingly, in Experiment 2, the stimuli were specifically aligned with the science curriculum in the host school system. They were generated from lessons that had not yet been presented in the classroom, ensuring the novelty of the information. The result was clear—open-ended self-derivation performance did not differ between experiments. The between-experiment comparison demonstrated that self-derivation through integration is operational across a wide range of content, including that featured in school children's science curriculum. There was a between-experiment difference in forced-choice recognition of the integration and stem facts, which was lower in Experiment 2 relative to Experiment 1 (though the difference for integration facts did not reach statistical significance). This pattern actually worked against strong open-ended self-derivation performance in that stem-fact memory is related to self-derivation. Thus lower levels of stem-fact memory would be expected to depress self-derivation. The fact that self-derivation performance was comparable across experiments is thus a testament to the robustness of self-derivation of new factual knowledge through memory integration even in the classroom, over science material.

The third primary purpose of the present research was to test whether children remember not only the stem facts explicitly taught to them in the context of story passages, but also other details of the explicit learning episodes (e.g., Butler et al., 2012). Studies of self-derivation through integration routinely test memory for stem facts (e.g., Bauer & Larkina, 2017; Bauer & San Souci, 2010). However, prior to the present research, there had been no tests of memory for the other information featured in the passages that are the vehicle for stem-fact presentation. The results of Experiment 2 were clear—one week after experience of the learning episodes, children accepted as “old” verbatim statements from the passages and rejected as “new” statements that were plausible but were not represented in the passages in any form. In both cases, performance was reliably different from that which would be expected by chance. On the basis of these patterns, we may conclude that children remember not only stem facts (tested in prior research) but also other of the information presented in the learning episodes (see also Butler et al., 2012).

The question of whether children retain information about the episodes in which stem facts are embedded raises the issue of whether the process of cross-episode integration extends beyond the stem facts to the episodic information surrounding them. Consider that the process of self-derivation depends on integration of stem facts, which results in an integrated fact representation. From the perspective of the child participant, there is no difference between the stem facts in the stories and the other information contained therein. Thus logically, stem facts and non-stem information are equally good targets for integration. Contrary to the assumption that non-stem information might be integrated, one week after experience of the learning episodes, children rejected statements that were hybrids of the two passages in a pair, created through cross-episode integration. Their levels of acceptance of the hybrid facts as “old” was reliably below chance. The same levels of performance were observed in high and low surface-similarity conditions. These patterns are contrary to those expected had the episodes been integrated with one another (i.e., integration should have been more likely in the high surface-similarity condition; an integrated representation should have led to acceptance of hybrid statements as “old”). Analyses of children's judgments of confidence in their endorsements of the statements indicated that children did not reject the hybrid statements out of confusion. Children had an intermediate level of confidence in their endorsements of hybrid statements relative to old and new statements. Finally, at the level of the individual trial, patterns of relation between confidence and endorsement were the same for all three statement types.

The possibility that integration might extend beyond the stem facts to the larger episode has important implications for memory of episodes of explicit instruction. If separate episodes of instruction concerning related topics are integrated with one another, then information about the source of the material could be lost to memory. This has especially important implications if some of the sources of information are more reliable than others. By 4 years of age, children take into account an informant's knowledge, expertise, and reliability (e.g., Birch, Vauthier, & Bloom, 2008; Robinson, Champion, & Mitchell, 1999). Yet if children do not maintain information about the source of information, these judgments are defeated, rendering children susceptible to misinformation. The present findings suggest that children do not integrate larger episodes of instruction, thus lessening the likelihood of this particular avenue to source confusion.

The possibility that source and other contextual details surrounding instructional episodes might be lost as a result of memory integration also bears on explanations for how episodes of instruction that are located in time and place give rise to semantic information that is timeless and placeless. That is, whereas episodes of instruction might be expected to result in specific episodic memories that are marked as to the *who, what, where, when, why,* and *how* of the experience, they are the source of semantic information that bears none of these features. The typical explanation for this transition from episodic to semantic memory is in terms of generalization over multiple episodes (e.g., Rogers & McClelland, 2004). Yet if the results of memory integration extend beyond the target facts upon which subsequent self-derivation depends, there is another means by which this transition is made—-through integration of separate learning episodes which then by default would no longer be tied to specific time and place. Speculation along this line was offered by Bauer and Jackson (2015) with respect to information derived from memory integration in adults. Here, we suggested that a similar fate could await the learning episodes that give rise to self-derived knowledge in children. The observation that the children in the present research apparently did not extend integration beyond the stem facts to the other information represented in the stimulus passages presents an interesting counterintuitive possibility, namely, that relative to adults, children may be slower to semanticize episodes of new learning owing to lack of integration. To our knowledge, this possibility has not been tested directly. Address of the question is a potential avenue for future research.

The question of whether full episodes are integrated with one another also sheds light on *when* memory integration takes place. In adults, the integration process seems to take place at encoding. This suggestion is based on unique patterns of ERP (Varga & Bauer, 2017a) and fMRI (e.g., Zeithamova et al., 2012) responses at the time of encoding for trials on which adults successfully derive correct responses versus trials on which they are unsuccessful. In contrast, in children, prior research suggests that memory integration does not occur without an explicit prompt to produce a response based on integration (Bauer et al., 2012, 2015; Varga & Bauer, 2013). In the self-derivation paradigm, an explicit prompt in the form of a question invokes the stem facts embedded in the story episodes (see Bauer & Varga, 2017, for discussion); it does not necessarily invoke the balance of information that surrounds the stem facts. Thus in effect, whereas there is a demand to integrate stem facts, there is no demand to integrate the other information in the stories. As such, if integration is found to extend beyond the stem facts, that suggests that integration took place without a prompt or demand, and thus most likely, at encoding. In the present research, the weight of the evidence suggests that for children, between-episode integration did not extend beyond the stem facts to the balance of the statements in the story passages—had it done so, children would have accepted hybrid statements as “old.” This pattern is consistent with prior observations that for children, memory integration does not take place at encoding, but only at test, in response to a demand (Bauer et al., 2012; Bauer et al., 2015; Varga & Bauer, 2013; see also Bauer, Dugan, Varga, & Riggins, 2019, for discussion). Even as we make this argument, we acknowledge that the evidence bearing on the timing of memory integration in children is indirect. Thus more definitive evidence is needed to fully address this question.

The results of the present research provide additional support for the contention that self-derivation of new factual knowledge through integration of separate yet related episodes of new learning is a valid model for accumulation of a knowledge base. Across experiments, the process has been shown to take place in the classroom as well as the laboratory, and over a range of content, including information derived from elementary science curricula. The products of the process—true but previously unknown facts—are retained in memory over at least one week, whether the facts were self-derived in the laboratory or the classroom. Moreover, children have accurate memory not only for the stem facts that are the targets of integration, but also other of the information conveyed in the learning episodes. The present research also reinforces a conclusion from prior research that suggests a restriction on the process of knowledge extension through memory integration, namely, that in childhood, integration may take place only in response to a prompt or demand. This conclusion serves as motivation for further research on the boundary conditions of self-derivation through integration, and on potential interventions to facilitate this important mechanism of learning.

Even as we highlight the significant contributions of the present research, we note some limitations of it. First, the participants were all 7–9 years of age. This age period was selected because we expected reasonably high levels of self-derivation performance and thus a strong test of memory for self-derived facts. Children of this age also could be expected to have relatively high levels of source memory, thus avoiding uninterpretable findings in Experiment 2. At the same time, the focus on a single age group constrains the generalizability of the findings and precludes test for developmental change in memory integration processes. Second, due to the requirement to present entire stories, and thus the length of the protocol, the story-passage paradigm used in the present research limits the number of trials that can be administered. This in turn limits the number of domains of information that can be sampled, as well as the number of trials available for the retention test. In future research, it would be desirable to use a variant on the self-derivation paradigm in which stem facts are conveyed in individual sentences rather than full stories, thus allowing for more delayed recall and recognition trials in the same amount of time (Bauer et al., 2016a; Esposito & Bauer, 2018).

A third limitation of the present research is that in Experiment 2, we obtained a relatively “coarse” assessment of children's endorsement of verbatim, new, and hybrid statements as “old.” That is, testing for recognition was forced-choice, with only two alternatives: children were required to endorse that they had heard the statement or had not heard the statement. We provided no avenue for them to tell us that they had heard a version of the statement, or had heard a portion of it, for example. This option is especially relevant in the case of hybrid statements which featured “old” material, combined in a “new” way. We adopted the dichotomous approach because the present research was the first test of the question. The logical starting point was to determine whether children in the target age range reliably indicated that material was old or new. In future research, it would be desirable to provide for more nuanced responses.

In conclusion, the present research provided clear address of all three of the questions that motivated it. We learned that children retain factual information they self-derive in the classroom, at least over delays of one week. We learned that self-derivation of new factual knowledge through integration of separate yet related episodes of new learning extends not only from the laboratory to the classroom, but to science content as well. We also learned that in addition to stem facts, children remember more of the information conveyed in the stimulus passages. The observation that children seemingly did not form integrated representations that extended beyond the stem facts also informs the question of when memory integration takes place—-consistent with prior research (Bauer et al., 2019; Bauer et al., 2012, 2015; Varga & Bauer, 2013), it seems that for children, memory integration may only take place in response to a demand or prompt. The findings simultaneously strengthen the argument that self-derivation through memory integration provides a valid model of accumulation of semantic knowledge, and sound a note of caution regarding this mechanism of learning, namely, that for children, it is not yet self-propelled, but must be kindled.

Acknowledgements

Support for this research was provided by NICHD HD67359, NSF BCS1528091, and IES R305A160240 to Patricia J. Bauer, and by Emory College of Arts and Sciences. The authors also extend their appreciation to the participating school system. The authors thank members of the Memory at Emory laboratory group for help at various stages of the research. They additionally extend their gratitude to the teachers who invited them into their classrooms, and to the children and parents who took part in the research.

Appendix C. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.learninstruc.2019.101271.

Appendix A

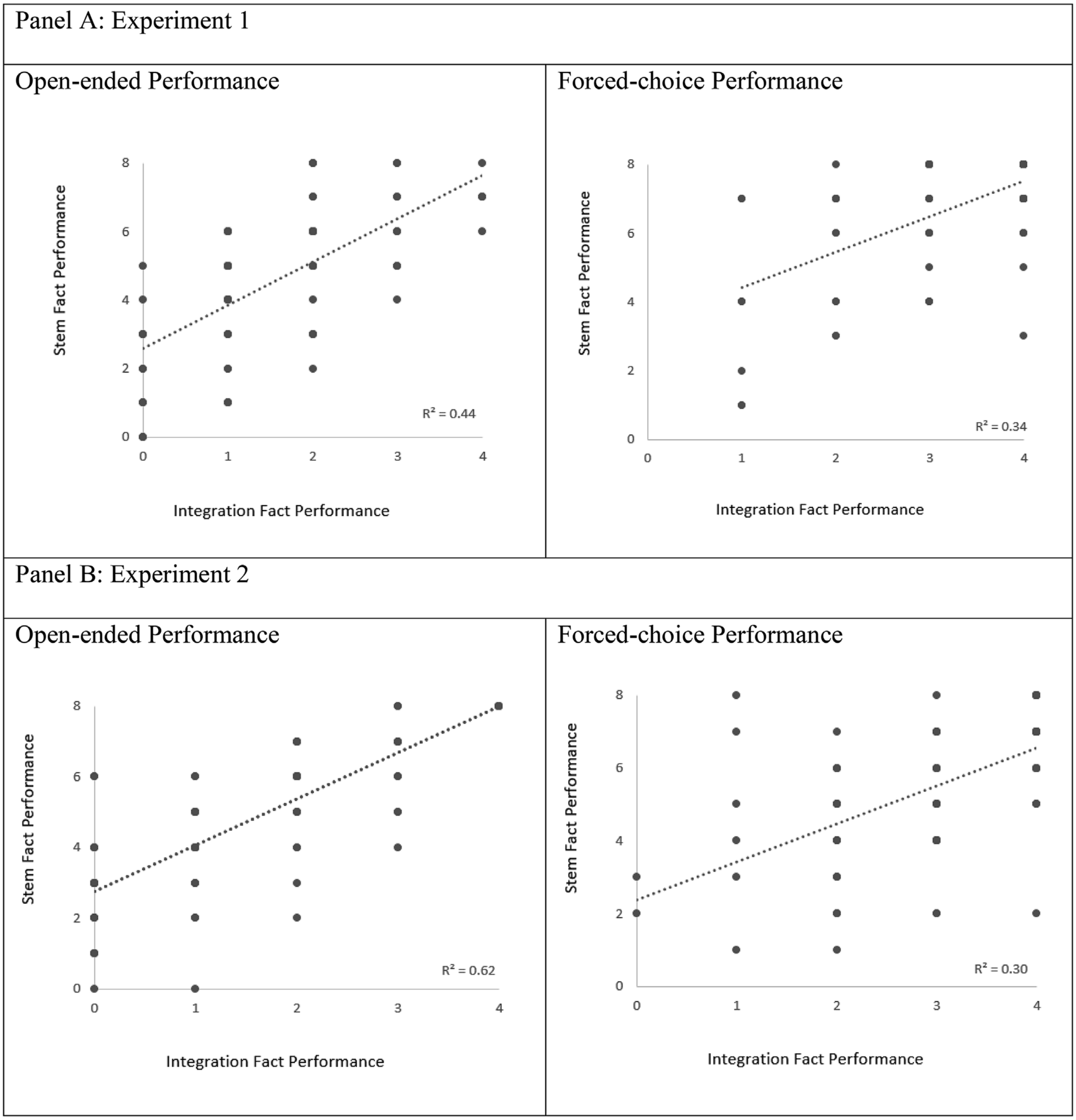

Figure A.1. Scatterplots depicting correlations between self-derivation of integration facts in open-ended and forced-choice testing and recall and recognition of stem facts in Experiment 1 (Panel A) and Experiment 2 (Panel B).1

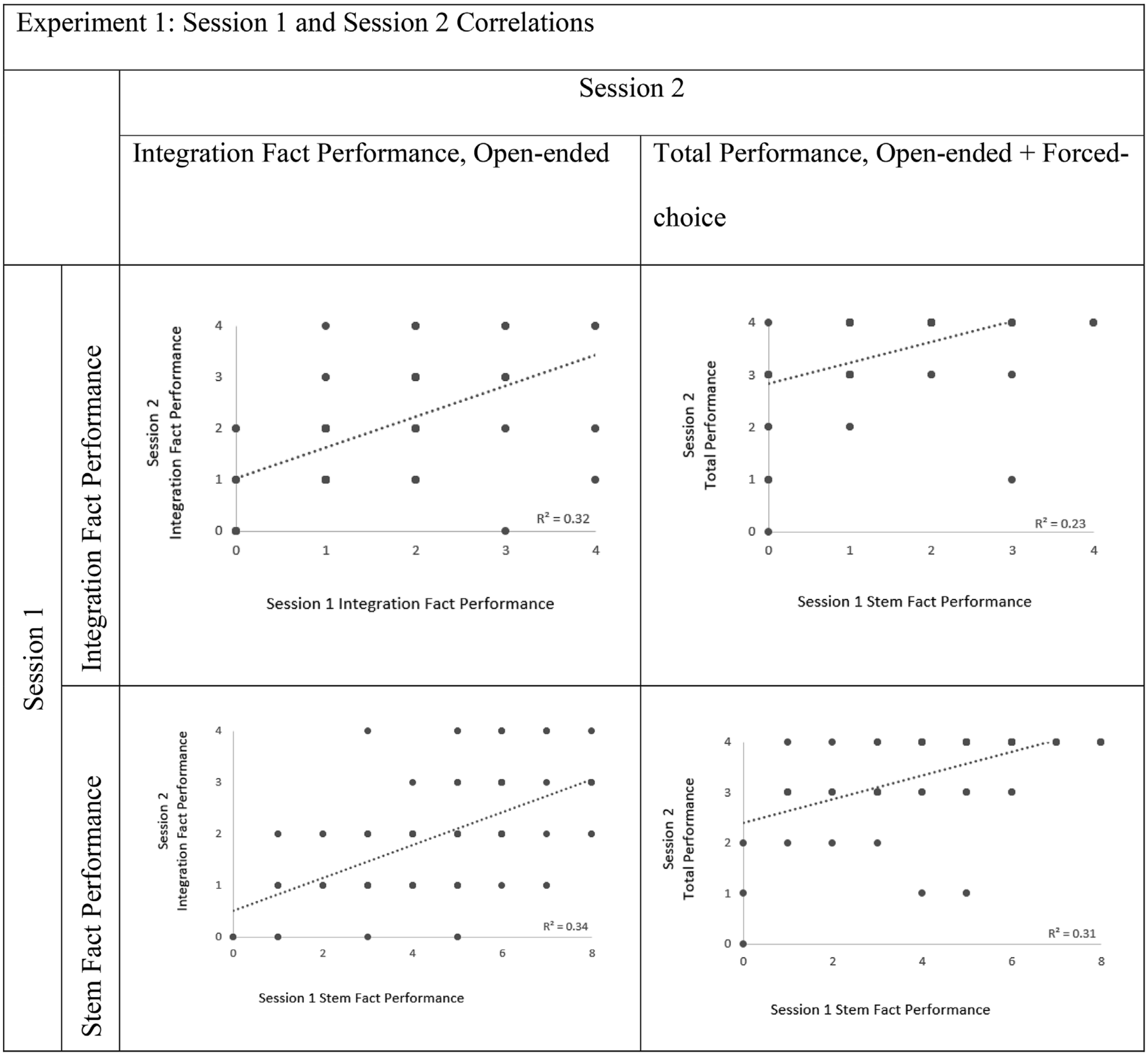

Figure A.2. Scatterplots depicting correlations between open-ended self-derivation of integration facts and recall of stem facts at Session 1 and recall and recognition of integration facts at Session 2 in Experiment 1.2

Appendix B

All multilevel model analyses were conducted with R (R Core Team, 2013) with the *nlme* package for non-linear mixed effect models. In each equation, the indices *i* and *t* are used to denote individual participants and trials, respectively, where in Level 1, the intercept, β_{0it} is defined as the expected mean endorsement of trial *t* of participant *i.* The error term, r_{it}, represents a unique effect associated with participant *i* (i.e., how much endorsement fluctuates within an individual across trials). The equations were used to test the null model, Model 1, and Model 2, for each statement type (Old, New, Hybrid). Goodness of fit was assessed with the Akaike's information criterion (AIC) and Schwarz's Bayesian criterion (BIC). Both are measures of model fit that correct for model complexity. Lower values indicate a better fitting model. Additionally, we evaluated the -2log-likelihood value (-2LL), or the deviance of the log-likelihood which is a measure of goodness of fit that is on a chi-square distribution. Models 1 and 2 both represented a significant improvement on the null model.

Null Model

A preliminary analysis was conducted to ensure that there was sufficient variability at Level 1 and Level 2 to warrant continuation with analyses. This preliminary analysis was a fully unconditional model (null model) in which no term other than the intercept was included at any level. The equations used to test the null models (Old, New, Hybrid) were:

Level 1: $\text{Endorsement}_{it} = \beta_{0it} + r_{it}$

Level 2: $\beta_{0i} = \gamma_{00} + u_{0i}$

- β_{0i} mean endorsement for participant *i*

- γ_{00} overall mean endorsement of the sample
- u_{0i} represented the degree to which individuals vary from the sample as a whole.

Model 1

The random coefficients regression models (Old, New, Hybrid) equations used to test whether self-derivation in Session 1 predicted endorsement of the statements in Session 2 were:

Level 1: $\text{Endorsement}_{it} = \beta_{0it} + \beta_{1it}(\text{self-derivation}) + r_{it}$

Level 2: $\beta_{0i} = \gamma_{00} + u_{0i}$

$\beta_{1i} = \gamma_{10} + u_{1i}$

The slope coefficient β_{1it} represented the associated change in endorsement associated with change in self-derivation performance. The individual intercept (β_{0i}) and slope (β_{1i}) become the outcome variables in the Level 2 equations, where γ_{00} represented the overall mean endorsement for the sample. Further,

- γ_{10} corresponded to the effect of self-derivation on endorsement.
- u_{0i} and u_{1i} represent the degree to which individuals vary from the sample as a whole.

Model 2

The random coefficients regression models (Old, New, Hybrid) equations used to test whether metacognitive judgements of endorsement predicted endorsement of the statements were:

Level 1: $_{\text{Endorsementit}} = \beta_{0it} + \beta_{1it}(\text{metacognitive judgement}) + r_{it}$

Level 2: $\beta_{0i} = \gamma_{00} + u_{0i}$

$\beta_{1i} = {}_{\gamma 10} + u_{1i}$

The slope coefficient β_{1it} represented the associated change in endorsement associated with change in metacognitive judgement. The individual intercept (β_{0i}) and slope (β_{1i}) become the outcome variables in the Level 2 equations, where γ_{00} represented the overall mean endorsement for the sample. Further,

- γ_{10} corresponded to the effect of metacognitive judgement on endorsement.
- u_{0i} and u_{1i} represent the degree to which individuals vary from the sample as a whole.

References

Bauer, P. J. (2012). Facilitating learning and memory in infants and young children: Mechanisms and methods. In S. L. Odom, E. P. Pungello, & N. Gardner-Neblett (Eds.). *Infants, toddlers, and families in poverty: Research implications for early child care* (pp. 49–74). New York, NY: The Guilford Press.

Bauer, P. J., Dugan, J. A., Varga, N. L., & Riggins, T. (2019). Relations between neural structures and children's self-derivation of new knowledge through memory integration. *Developmental Cognitive Neuroscience, 36*, 100611. https://doi.org/10.1016/j.dcn.2018.12.009.

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.

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*, 681–694. https://doi.org/10.1016/j.jecp.2011.10.005.

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.

Bauer, P. J., & San Souci, P. (2010). Going beyond the facts: Young children extend knowledge by integrating episodes. *Journal of Experimental Child Psychology, 107*, 452–465.

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

Bauer, P. J., Stewart, R., White, E. A., & Larkina, M. (2016b). A place for every event and every event in its place: Memory for locations and activities by 4-year-old children. *Journal of Cognition and Development, 17*(2), 244–263. https://doi.org/10.1080/15248372.2014.959521.

Bauer, P. J., & Varga, N. L. (2017). Similarity and deviation in event segmentation and memory integration. [Peer commentary on the paper "Event perception: Translations and applications" by L.L. Richmond, D.A. Gold, & J.M. Zacks]. *Journal of Applied Research in Memory and Cognition, 6*(2), 124–128.

Bauer, P. J., Varga, N. L., King, J. E., Nolen, A. M., & White, E. A. (2015). Semantic elaboration through integration: Hints both facilitate and inform the process. *Journal of Cognition and Development, 16*, 351–369. https://doi.org/10.1080/15248372.2013.849707.

Birch, S. A. J., Vauthier, S. A., & Bloom, P. (2008). Three- and four-year-olds spontaneously use others' past performance to guide their learning. *Cognition, 107*(3), 1018–1034. https://doi.org/10.1016/j.cognition.2007.12.008.

Bjork, R. A., & Richardson-Klavehn, A. (1989). On the puzzling relationship between environmental context and human memory. In C. Izawa (Ed.). *Current issues in cognitive processes* (pp. 313–344). Hillsdale, NJ: Erlbaum.

Bluck, S., Alea, N., Habermas, T., & Rubin, D. C. (2005). A TALE of three functions: Self-reported uses of autobiographic memory. *Social Cognition, 23*, 91–117.

Butler, A. C., Dennis, N. A., & Marsh, E. J. (2012). Inferring facts from fiction: Reading correct and incorrect information affects memory for related information. *Memory, 20*(5), 487–498.

Cycowicz, Y. M., Friedman, D., Snodgrass, J. G., & Duff, M. (2001). Recognition and source memory for pictures in children and adults. *Neuropsychologia, 39*, 255–267.

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.

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.

Esposito, A. G., & Bauer, P. J. (2019). Self-derivation through memory integration under low surface-similarity conditions: The case of multiple languages. *Journal of Experimental Child Psychology, 187*, 104661.

Fazio, L. K., & Marsh, E. J. (2019). Retrieval-based learning in children. *Current Directions in Psychological Science, 28*(2), 111–116.

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

Gentner, D. (1989). The mechanisms of analogical learning. In S. Vosniadou, & A. Ortony (Eds.). *Similarity and analogical reasoning* (pp. 199–241). New York, NY, US: Cambridge University Press.

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

Herz, R. (1997). The effects of cue distinctiveness on odor-based context-dependent memory. *Memory & Cognition, 25*(3), 375–380.

Jacoby, L. L., & Craik, F. I. M. (1979). Effects of elaboration of processing at encoding and retrieval: Trace distinctiveness and recovery of initial context. In L. S. Cermak, & F. I. M. Craik (Eds.). *Levels of processing in human memory* (pp. 1–22). Hillsdale, NJ: Erlbaum.

McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review, 99*(3), 440–466.
Newcombe, N. S., Ambady, N., Eccles, J. S., Gomez, D., Linn, M., Miller, K., et al. (2009). Psychology's role in mathematics and science education. *American Psychologist, 64*, 538–550.
Nezlek, J. B. (2001). Multilevel random coefficient analyses of event- and interval-contingent data in social and personality psychology research. *Personality and Social Psychology Bulletin, 27*, 771–785.
Pillemer, D. B. (2003). Directive functions of autobiographical memory: The guiding power of the specific episode. *Memory, 11*(2), 193–202.
Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*, *Vol. 1*. Thousand Oaks, CA: Sage Publications.
Riggins, T. (2014). Longitudinal investigation of source memory reveals different developmental trajectories for item memory and binding. *Developmental Psychology, 50*, 449–459. https://doi.org/10.1037/a0033622.
Robinson, E. J., Champion, H., & Mitchell, P. (1999). Children's ability to infer utterance veracity from speaker informedness. *Developmental Psychology, 35*, 535–546. https://doi.org/10.1037/0012-1649.35.2.535.
Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach.* Cambridge, MA: MIT Press.
Siegler, R. S. (1989). Mechanisms of cognitive development. *Annual Review of Psychology, 40*(1), 353–379.
Smith, S. M. (1988). Environmental context-dependent memory. In G. M. Davies, & D. M. Thomson (Eds.). *Memory in context: Context in memory*. Toronto: Wiley.
Tulving, E. (1972). Episodic and semantic memory. In E. Tulving, & W. Donaldson (Eds.). *Organization of memory* (pp. 381–403). New York: Academic Press.
Tulving, E. (1983). *Elements of episodic memory.* Oxford: Oxford University Press.
Varga, N. L., & Bauer, P. J. (2013). Effects of delay on 6-year-old children's self-generation and retention of knowledge through integration. *Journal of Experimental Child Psychology, 115*, 326–341. https://doi.org/10.1016/j.jecp.2013.01.008.
Varga, N. L., & Bauer, P. J. (2017a). Young adults self-derive and retain new factual knowledge through memory integration. *Memory & Cognition, 45*(6), 1014–1027.
Varga, N. L., & Bauer, P. J. (2017b). Using event-related potentials to inform the neurocognitive processes underlying knowledge extension through memory integration. *Journal of Cognitive Neuroscience, 29*(11), 1932–1949.
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.
Varga, N. L., Stewart, R. A., & Bauer, P. J. (2016). Integrating across episodes: Investigating the long-term accessibility of self-derived knowledge in 4-year-old children. *Journal of Experimental Child Psychology, 145*, 48–63. https://doi.org/10.1016/j.jecp.2015.11.015.
Waters, T. E. A., Bauer, P. J., & Fivush, R. (2014). Autobiographical memory functions served by multiple event types. *Applied Cognitive Psychology, 28*, 185–195.
Wright, D. B. (1998). Modeling clustered data in autobiographical memory research: The multilevel approach. *Applied Cognitive Psychology, 12*, 339–357.
Zeithamova, D., Dominick, A. L., & Preston, A. R. (2012). Hippocampal and ventral medial prefrontal activation during retrieval-mediated learning supports novel inference. *Neuron, 75*(1), 168–179.