Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____   _____
Jessica Ann Dugan   Date

Metacognition in Self-derivation across Development

By

Jessica A. Dugan
Doctor of Philosophy

Psychology

_____
Patricia J. Bauer, Ph.D.
Advisor

_____
Robert Hampton, Ph.D.
Committee Member

_____
Joseph Manns, Ph.D.
Committee Member

_____
Janet Metcalfe, Ph.D.
Committee Member

_____
Phillip Wolff, Ph.D.
Committee Member

Accepted:

_____
Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

_____
Date

Metacognition in Self-derivation across Development

By

Jessica A. Dugan

B.S. (Highest Honors), The Honors College at the College of Charleston, 2014

M.A., Emory University, 2016

Advisor: Patricia J. Bauer, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Psychology
2020

Abstract

Metacognition in Self-derivation across Development

By Jessica A. Dugan

The acquisition of knowledge is a crucial task across the lifespan. A pervasive means by which knowledge is acquired is self-derivation through the integration of separate yet related episodes of new learning. Though successful self-derivation of new knowledge has been observed in individuals across development, from early childhood and into adulthood, there is considerable age-related and individual variability in the success one experiences. Metacognition, "thinking about thinking," may be related to age-related and individual variability observed in self-derivation performance. This is because metacognition may support learners monitoring their knowledge for the facts necessary to self-derive new knowledge and it may also facilitate the steps necessary to successfully integrate and derive new knowledge. In 3 studies, we investigated the relation between self-derivation and metacognition. In the first direct examination of adults' self-derivation and metacognition, adults made judgments about how well they believed they learned facts necessary to self-derive new knowledge. In Study 1, adults were more likely to self-derive when they made a judgment about one or both of the necessary facts than when they did not make a judgment. In Study 2, a cross-sectional examination of children's metacognition about their self-derivation, there was not significant age-related change in the relation between children's metacognitive accuracy and their self-derivation performance. However, the accuracy of confidence judgments made after answering self-derivation questions was significantly related to self-derivation performance suggesting that children can successfully monitor their self-derivation. Study 3 was a 3-year longitudinal investigation of children's self-derivation and metacognition. Results revealed significant stability in self-derivation performance from year to year. Children's metacognition about their self-derivation did not begin to stabilize until Year 3 when children's predictions about their performance were related to the previous year's predictions. Taken together, these data suggest that metacognition about self-derivation undergoes significant development in late childhood and that children can glean some information about their self-derivation performance through metacognitive monitoring. More broadly, these results suggest that metacognition likely needs to be actively engaged to be most effective in the context of self-derivation across development.

Metacognition in Self-derivation across Development

By

Jessica A. Dugan

B.S. (Highest Honors), The Honors College at the College of Charleston, 2014

M.A., Emory University, 2016

Advisor: Patricia J. Bauer, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Psychology
2020

Acknowledgements

There are many, many people who supported me and this work. I would first like to thank my mentor and advisor Patricia J. Bauer for giving me the opportunity to earn my PhD and become a scientist. I am both forever changed and forever grateful for the chance Patricia took on me. A significant portion of my graduate school years and of the dissertation work are a result of Alena Esposito bringing me on as a collaborator on a wonderful longitudinal project. It is special to have a great colleague and friend in Alena. Thanks are also due to Nicole Varga for going on countless runs with me and making sure I never abandoned my dreams. Friend, you are a brilliant scientist and a diehard friend.

There are several students and lab members who contributed directly to the research discussed below. I would like to thank Reilly Allison for assistance with stimulus generation and Katie Lee for her help with data collection and other aspects of the research. Emily Beattie, Ben Gilbert, Alissa Miller, and Emma Simpson each devoted hundreds of hours checking, and organizing, and rechecking the data. Credible and informative science would not prevail without the dedication you provided to the cause. Great thanks are also due to my lab family over the years and those who provided care and support throughout the dissertation writing process. I cannot help but smile when I think of the pep squad you were for me.

I would also like to thank my wife, Lucy Hamer, for her support throughout the completion of the dissertation. Having your support means more than I can put into words. Lastly, I also extend my gratitude to the individuals and families who took part in the research, without whom this work would not have been possible.

Table of Contents

List of Tables

Study 1

Study 2

Study 3

Table 7. *Multilevel models of self-derivation performance as predicted by metacognitive measures each year: Model by year (Panel A), feeling of knowing as a single predictor (Panel B), feelings of knowing and confidence judgments as predictors of FC self-derivation (Panel C).*

Table 8. *Regression analysis summary for Year 1 self-derivation and metacognition predicting self-derivation at Year 2.*

Table 9. *Regression analysis summary for Year 1 self-derivation an metacognition predicting self-derivation at Year 3.*

Table 10. *Regression analysis summary for Year 2 self-derivation and metacognition predicting self-derivation at Year 3.*

List of Figures

General Introduction

For a moment, imagine being a contestant on the game show *Jeopardy!* At the beginning of the first round host Alex Trebek reads the categories aloud. One of them is your area of expertise. When it is your turn to select the prompt you immediately go for the highest value available in that category, confident that you will answer correctly. Later, when the last category is the one you have been avoiding the whole round, you try to play it safe by choosing the lowest value item. Though most people will never play as *Jeopardy!* contestants on television, in this example there are two components of cognition relevant to daily life: semantic knowledge about the world, including topics such as literature, music, history, biology, physics, as well as knowledge about one's self – specifically one's own cognition. Knowledge about the world is gained through both direct experiences and productive processes that allow for extension of information learned directly. Self-derivation through memory integration, the focal productive process of the present research, is an essential process that contributes to the accumulation of knowledge about the world across development (Bauer & Varga, 2017; Varga, Esposito, & Bauer, 2019). Metacognition, one's thoughts about her cognition, plays an important role in informing and guiding one's learning and memory (Flavell, 1979; Nelson & Narens, 1990). Yet it is currently unknown what impact metacognition may have on self-derivation of new knowledge in children or adults.

This 3-part dissertation is the first direct investigation of metacognition and self-derivation. Previous research on metacognition and productive processes has been limited to specific reasoning processes, such as transitive inference (Smith & Squire, 2005). Study 1 is an investigation of the impact of making metacognitive judgments when encoding related fact pairs upon adults' subsequent self-derivation. Study 2 is a cross-sectional examination of the relation between children's metacognition and self-derivation across grades 3 through 5. Finally, Study 3

is a 3-year longitudinal investigation of children's metacognition about their self-derivation, spanning grades 3 through 5. Together, these investigations of metacognition as it relates to self-derivation make a novel contribution to the present understanding of how one's knowledge of their cognition relates to the internal cognitive processes that are productive processes.

Learning and Self-derivation

Knowledge accumulation is an essential task across development, whether it be a prerequisite for participation on *Jeopardy!*, the development that allows advancement into the next school grade, or the expertise that allows for success in one's career. The materials that build a knowledge base are learned through direct experiences such as a classroom lecture, an educational text, or events like a trip to the zoo. Importantly, learning does not occur only in these direct ways; learning also occurs through productive processes that allow one to go beyond the information provided to learn something new. Such processes include, but are not limited to, analogy (Gentner & Smith, 2012), transitive inference (Smith & Squire, 2005), associative inference (Zeithamova, Dominick, & Preston, 2012), and the focus of the present work – self-derivation (Bauer & San Souci, 2010; Bauer & Varga, 2017). Self-derivation is the process by which new factual knowledge is derived from the integration of separate but related episodes of new learning. In our laboratory, we engineer self-derivation by presenting participants with separate but related "stem" facts that can be integrated with one another and then posing questions that can be addressed by self-deriving new knowledge over the integrated fact representation (Bauer & San Souci, 2010; Bauer & Varga, 2017). To illustrate, while watching the evening news you may learn that *some beer is produced using isinglass*. Some time later you learn *isinglass is an adhesive found in fish*. Integration of these facts in memory allows for derivation of the new understanding that *some beer is produced using an adhesive found in fish*.

Though this example is likely salient to those who enjoy the craft of fermentation, self-derivation is quite pervasive in that it is observed across development (Bauer & San Souci, 2010; Varga & Bauer, 2017) and over various media (Dugan & Bauer, 2020).

Self-derivation is pervasive across development. It has been observed in children as young as four years-old (Bauer & San Souci, 2010; Varga & Bauer, 2013), across the elementary school years (Bauer, Blue, Xu, and Esposito, 2016; Bauer & Larkina, 2017), and in college students (Bauer & Jackson, 2015; Dugan & Bauer, 2020; Varga & Bauer, 2017a; 2017b). Substantial increases in derivation of new knowledge occur between 4- and 8-years (Bauer & Larkina, 2017). To illustrate, Bauer and Larkina (2017) read 4-, 6-, and 8-year-olds stories that contained novel but true facts that could be integrated to self-derive new understandings. For example, in one story children learned *the largest desert in the world is the Sahara.* Later, in another story children learned *the largest desert in the world is located in Africa.* They then are asked questions that probe for extension beyond the individual stories, through integration of them. Only 13% of 4-year-olds successfully self-derived the new knowledge that *the Sahara is located in Africa* whereas 50% of 6-year-olds and 75% of 8-year-olds did so. This provides strong evidence that self-derivation through integration increases with age across childhood.

In addition to age-related change there are also individual differences that contribute to variability in self-derivation. Varga and colleagues (2017a; 2019) conducted experiments with both children and adults. Their sample of adults ranged from 3% to 93% correct in self-derivation performance, with 50% correct on average (Varga & Bauer, 2017a); children ranged from 0% to 100% correct, averaging 38% correct on self-derivation (Varga et al., 2019). Varga, Esposito, and Bauer (2019) found that working memory and verbal comprehension are unique contributors to the observed variability in self-derivation of new factual knowledge among

adults. More specifically, they found that verbal comprehension and working memory were the only significant predictors of self-derivation performance from amongst a battery of cognitive measures. Only verbal comprehension accounted for a significant amount of the variability in children's performance. This work emphasizes the presence of individual differences both within and across age groups, suggesting that the variability is systematic (i.e., it relates to other cognitive measures is meaningful ways), and thus the need to further investigate the variables underlying these differences.

It is difficult to overstate the importance of self-derivation as a means of knowledge accumulation. Self-derivation is in the family of productive, inferential processes, including induction and deduction, considered significant mechanisms of building a knowledge base (Goswami, 2011). In productive processes like transitive inference and associative inference, training and/or explicit instruction are usually necessary for participants to successfully learn new information. In contrast, participants who successfully self-derive new knowledge first learn separate but related facts distributed over a learning episode but they never receive direct instruction or training related to the task (Bauer & Larkina, 2017; Varga & Bauer, 2017a). What might be most relevant about self-derivation as a mechanism of constructing knowledge is that it readily accommodates the way we often learn information – distributed over time. Because learning is distributed across episodes and sources, the need for integration and subsequent self-derivation is pervasive.

Self-derivation through memory integration is ecologically relevant to life outside the laboratory. Content learned as stem facts and the knowledge that is subsequently gained in the process of self-derivation, is factual and often aligned with materials learned in students' coursework (Esposito & Bauer, 2017). Recent self-derivation work has even been extended to

materials available for the consequential decisions of everyday life such as what prescriptions drugs are appropriate for consumption (Dugan & Bauer, 2020). Though some studies of other productive processes employ rich, meaning-laden stimuli (Gick & Holyoak, 1980; 1983) many such examinations, including those testing analogy and associative inference, typically operate over only 2 to 4 images or words. In the cases of transitive (Lazareva & Wasserman, 2010) and associative (Preston & Eichenbaum, 2013; Zeithamova et al., 2012) inference, the images or words are arbitrarily paired. Knowledge gained through self-derivation is rapidly incorporated into the knowledge base as evidenced by adults' neural responses to integration facts (Bauer & Jackson, 2015). Not only is self-derivation a practical means of generating new knowledge for its flexibility across development and content, the process and products of self-derivation are effective for long-term use. Self-derived knowledge is retained for at least one week by both children (Dugan & Bauer, 2020; Varga & Bauer, 2013) and adults (Varga & Bauer, 2017a). Self-derivation performance is also related to concurrent and later academic achievement (Varga, Esposito, & Bauer, 2019). In sum, self-derivation is the focus of the present work because it has many desirable features (e.g., factual stimuli, pervasive demand for integration, training not required) that have not been evidenced of other productive processes and because self-derivation has been explored for its potential role in constructing a knowledge base, which has not been a primary aim of other productive processes.

Possible Role of Metacognition in Self-derivation

Sources of variability, like verbal comprehension and working memory, likely contribute to differences in self-derivation performance because they directly impact one (or more) of the subprocesses of self-derivation. Metacognition may likewise impact self-derivation at any one of these subprocesses. Metacognition is uniquely suited to examination of the variability in self-

derivation performance and what individuals may understand of this process because the knowledge that arises from self-derivation is created in the mind of the learner rather than experienced directly. To better understand both the age-related and individual variability in self-derivation performance, we have developed a 5-step process model: ERISS (Bauer & Varga, 2017). Utilizing a previous fact-pair example, the five steps that metacognition may impact are as follows. First, an initial fact is Encoded (the "E" in ERISS) *the largest desert in the world is the Sahara*. This fact is then Reactivated upon encoding of a second, related fact *the largest desert in the world is in Africa*. These two facts are then Integrated in memory with each other based on the shared information *the largest desert*, as well as with any prior knowledge of deserts. Upon a demand – *where is the Sahara desert?* – relevant information necessary to meet the demand is Selected from among the knowledge base about deserts. The new knowledge that *the Sahara desert is in Africa* is then Self-derived.

Evidence suggests that variability in successful engagement of the steps of ERISS contributes to both age-related change and individual variability in successful self-derivation (see Bauer, Cronin-Golomb, Porter, Jaganjac, & Miller, (in press) and Bauer & Larkina, 2017, for discussion). There is reason to speculate that metacognitive processes may also play a role. Generally speaking, metacognition is cognition about one's cognition or "thinking about thinking" (Flavell, 1979). More specifically, metacognition is the combination of monitoring and control processes that support the development of knowledge about our semantic memory, learning strategies, and the ways in which these interact to contribute to the knowledge base (Nelson & Narens, 1990). For example, selecting a high value question on *Jeopardy!* when lower alternatives remain on the board suggests reflection upon the contents and confidence in one's memory. Similar cues about learning and memory may be applied in the context of self-

derivation. Metacognitive monitoring involves what we know about our cognition. This knowledge comes from observing and making judgments about how well an item has been learned, for example. It includes the feelings experienced when acquiring knowledge or attempting to retrieve it. Flavell (1979) referred to such feelings as "metacognitive experiences" because they involve a subjective feeling of knowing something with certainty, being confused, or lacking confidence in one's own knowledge. Metacognitive control refers to the many steps we take to modify our cognition including looking for strategies that will improve learning outcomes, initiation of a strategy to test its efficacy, and eventually termination of that strategy when one's learning goal has been realized or when it is clear that attempt at regulation was unsuccessful. Monitoring is evaluated through elicited metacognitive judgments, whereas attempts to understand control have included measures of study duration, strategy selection, and item selection.

How might the metacognitive processes of monitoring and control relate to self-derivation of new knowledge through memory integration? Participants need not be highly confident in their performance to be accurate. Rather, accurate metacognitive monitoring involves being appropriately confident in one's own knowledge and being appropriately uncertain when one does not have knowledge necessary for successfully completing a task. Of critical importance for present purposes is that because the subprocesses of self-derivation through integration unfold over time, there is ample opportunity for the metacognitive processes of monitoring and control to influence self-derivation performance. Each of the following hypotheses is based on the assumption that a participant's metacognition is accurate and that metacognition works toward successful self-derivation.

If a participant accurately monitors their learning, then learning of the individual stem facts may be increased owing to acute attention to and perhaps even elaborative encoding of the facts that elicited higher confidence. This likely results in the individual being more likely to self-derive on more trials because knowledge of stem facts is necessary for self-derivation of new knowledge. Actively monitoring learning may also contribute to participants' awareness of two key aspects of self-derivation through integration. As participants encode separate but related stem facts distributed throughout a set of facts about different topics, participants may become keenly aware of the related nature of the pairs of facts. They then might begin to identify the information that is shared within a pair of stem facts. Monitoring may direct attention after encoding Stem 2 so that only relevant and related information – Stem 1 – is reactivated. Such knowledge could also allow a participant to deliberately engage in Integration. Metacognition may control Selection of necessary, relevant facts, from a corpus of knowledge, in service of deriving new understandings. Metacognitive monitoring may also contribute to greater success at Selection as the elaborative encoding supported by monitoring one's learning may "tag" relevant stem facts differently from other knowledge one might have about a topic. Finally, there remains the possibility that even if encoding, reactivation, integration, or selection are not directly facilitated by successful metacognitive monitoring and control on the level of the individual trial, learners may extract the general rule that in this task, some information is related to other learned information over time and can be put together to form additional understandings, ultimately resulting in more new knowledge than those who are unaware of the relational structure present within the fact pairs

Using Metacognition as "Leverage" to Understand Learning

As illustrated above, metacognition is a resource for learners to use as they study or acquire new knowledge. Importantly, it is also a tool researchers can employ to better understand cognition. For example, the judgment of learning is one particularly good measure of metacognition. Judgments of learnings (JOLs) are prospective judgments of how well one believes she or he has learned either individual items or a set of items. Judgments of learning can be made immediately after studying items or following a delay. For example, in a metamemory study with paired associates a participant will see a pair of words, *apple-blanket*, on a computer screen. Once the pair disappears, participants are then asked "What are the chances that you will recall the second word when presented with the first word?" (Koriat, 1997). Participants indicate their responses as percentages between 0% and 100%. Gamma correlations, an indicator of metacognitive accuracy, are calculated by correlating participants' responses with their subsequent recall of test items. Delayed JOLs are more strongly related to performance than immediate JOLs yielding mean gamma correlations of .90 and .38, respectively (Nelson & Dunlosky, 1991).

Interestingly, accuracy of item-by-item JOLs improves with repeated studying (Koriat, 1997). Koriat (1997) posits that repeated studying improves JOL accuracy because with continued study of the same items participants begin relying on internal cues about memory for the stimuli, suggesting that they attend to how easy or hard it is to remember those items. Importantly, evidence from several experiments with paired associate learning demonstrate that adults use mnemonic cues as the basis of the JOLs, rather than intrinsic or extrinsic cues (Koriat, 1997). These findings highlight the utility of JOLs as windows into participants' learning. Judgments of learning are the most widely-used judgments (Koriat, 2007) and thus, a notable candidate for better understanding the process of self-derivation. According to Son and Metcalfe

(2005), judgments of learning occur as one attends to information provided about the memory and then after the retrieval attempt.

Another index of metacognitive monitoring is the retrospective confidence judgment. Confidence judgments (CJs) are elicited after learning, at test. After providing an answer at test, participants then nominate how certain they are that they have answered correctly. Confidence judgments such as these are generally thought to be based on the retrieval experience of testing (Nelson & Narens, 1990; Koriat, 2012), reflecting a veridical measure of participants' metacognitive monitoring. These judgments are also relatively consistent within individuals suggesting that individuals engage in a similar process of monitoring across tasks (Ais, Zylberberg, Barttfeld, & Sigman, 2016) and making them amenable to new applications. Metacognitive judgments like JOLs and CJs have the potential to provide insight into participants' understanding of self-derivation. A significant relation between the accuracy of such judgments and self-derivation performance would suggest that participants, whether adults or children, utilize their knowledge of their cognition to guide self-derivation performance.

Leveraging Metacognition over Development

Evidence suggests that children's metacognitive monitoring judgments generally follow the same patterns as adults' and that accuracy of these judgments increases with development. Like adults, children show greater accuracy in judging how well they have learned an item when the judgment is made following a delay of even two minutes. In contrast to adults, children tend to be overconfident in their item-by-item predictions of ease-of-learning (Visé & Schneider, 2000). Current evidence on children's ability to predict future performance on test items is mixed but suggests that the accuracy of these feeling-of-knowing judgments is relatively low overall, with average gamma correlations below .25. (Lockl & Schneider, 2002). In general, accuracy of

each kind of judgment improves in subtle increments across development (Schneider & Lockl, 2008). Some research does suggest that monitoring and control develop on different developmental trajectories, but further investigation into mapping these trajectories is warranted by the relative sparsity of data on control processes over development (Roebers et al., 2019; Roebers, Krebs, & Roderer, 2014; Schneider & Lockl, 2002; 2008).

The overall pattern of relations between memory performance and metacognitive judgments also is similar for children and adults, yet there are notable differences in relations over development. Overall, adults exhibit greater relative accuracy in their metacognitive judgments. As such, metacognitive accuracy may contribute to age-related variability. Alternatively, metacognitive accuracy may impact self-derivation performance such that individuals with well-calibrated systems are more successful than those who cannot use their metacognition to direct their self-derivation performance. If a relation between self-derivation and metacognition emerges over development, examining metacognition about self-derivation across children and adults over time will help reveal that relation. In summary, metacognition may contribute to variability observed in performance due to either, or both, age-related change or individual differences.

Empirical Tests of Relations between Self-derivation and Metacognition

To date there has been no work directly testing relations between self-derivation through integration and metacognition. There is suggestive evidence of such relations, however. Evidence from adults who engage in multiple blocks presenting the opportunity to self-derive (i.e., encode—test; encode—test; etc.) shows that performance improves from the first block to subsequent blocks, despite participants never receiving instructions on how to perform successfully (Menkes, Esposito, & Bauer, 2017). In a separate test of adults' self-derivation,

Varga and Bauer (2017) consulted participants' knowledge of relatedness between pairs of facts. They found a moderately strong relation between knowledge of relatedness and successful self-derivation, $r = .44$. Furthermore, participants who were unsuccessful on a subset of trials but reported being aware of the opportunity to integrate exhibited longer response times specifically for the unsuccessful trials. This suggests participants who understood integration was the goal searched their memories longer for a correct response when they were wrong (Varga & Bauer, 2017). Taken together, these data highlight that not only is there individual variability in self-derivation performance but that this variability is likely related, at least in part, to differential metacognitive monitoring and control as self-derivation unfolds.

The present research is a series of investigations examining these questions in both children and adults. In adults, we tackle the basic question of whether metacognition impacts self-derivation. In children, we address both change over time between and within individuals over a three-year period. These are the first direct investigations of metacognition and self-derivation as they may work together.

In Study 1, adults learned a series of fact pairs. For a subset of the pairs, participants provided a judgment of learning (JOL) for one or both of the necessary stem facts. We tested whether a judgment of learning impacts subsequent self-derivation. We hypothesized that making a judgment of learning at encoding would increase self-derivation performance by increasing memory for the individual stem facts. This is important because there is limited but mixed evidence on the impact of metacognitive judgments on subsequent memory suggesting that making a judgment does influence memory (Dougherty et al., 2005; 2018). Furthermore, if making a metacognitive judgment does improve self-derivation performance, then that provides support of metacognition as a source of individual variability.

In Study 2, elementary school students in grades 3, 4, and 5 learned pairs of facts in their home classrooms. At test for self-derivation, children provided feelings-of-knowing (FOK) judgments that serve as a prediction of future performance on forced choice self-derivation questions. We examined children's metacognitive accuracy of their FOKs as well as the accuracy of retrospective confidence judgments given after responding to forced choice questions. We also examined age-related differences across the three grades. We expected that self-derivation performance would be higher in later grades (at older ages) as has been observed prior research (Bauer & Larkina, 2017; Esposito & Bauer, 2017). Based on previous research on children's metacognition (Lockl & Schneider, 2002; Roebers et al., 2019), we also expected that metacognitive accuracy would increase across the three grade levels. If self-derivation and metacognition are not related in grade 3 but are related in grade 4 or 5, then this could suggest that development allows them to work in concert, in service of successful self-derivation.

Study 3 was a longitudinal investigation of a cohort of elementary students who engaged in the same experiment as in Study 2. This cohort of children participated for 3 consecutive years, starting in Grade 3 and ending in Grade 5. With these data, we examined age-related change within the same individuals over time. We expected to observe change in both self-derivation and metacognitive measures over 3 years because prior research reflects that these two components independently develop over childhood (Bauer & Larkina, 2017; Roebers et al., 2019). Observing the emergence of a significant relation between these two variables would suggest that metacognition contributes to the age-related increases in self-derivation performance. It could also indicate metacognition is a source of individual variability. Importantly, the longitudinal design provides a stronger test of development than cross-sectional work because the same children are observed over time.

Study 1

Metacognitive Judgments Improve Adults' Self-derivation through Memory Integration

Jessica A. Dugan & Patricia J. Bauer

Emory University

Metacognitive Judgments Improve Adults' Self-derivation through Memory Integration

Compiling a knowledge base is a fundamental task of life. Indeed, whether or not knowledge is intentionally sought, humans are constantly engaged in opportunities for learning. Knowledge can be gained through direct experiences including lectures, books, and even commercials. Learning and memory are also remarkable for the capacity to extend beyond separate direct experiences to accrue knowledge across time, contexts, and media via productive processes such as self-derivation. A striking feature of self-derivation is the range of variability adults exhibit in self-derivation performance, from virtually floor to virtually ceiling levels (Varga & Bauer, 2017a). Self-derivation is also significantly related to academic success (Varga, Esposito, & Bauer, 2020), and as such it is critical to understand the factors contributing the observed individual variability. A logical candidate contributor to individual variability is metacognition. Metacognition can guide an individual toward successful learning outcomes by, for example, assisting one in predicting future performance on a task and in attending to items that are related to each other or require further study for successful performance. This postulated connection between self-derivation and metacognition has not yet been the focus of scientific inquiry. In the present research, we addressed this gap by examining both the metacognitive judgments adults made as they learned pairs of integrable facts and the relation between those judgments and subsequent self-derivation performance.

Learning opportunities are abundant and varied in where they occur, what they are about, and how they are presented. Direct opportunities for gaining knowledge are those easily recognizable as occurring in a classroom, the pages of a book, or the exhibits at a zoo. Importantly, learning also occurs via productive processes that allow one to go beyond information gained through direct experiences to learn something new. Productive processes include, to name a few, analogy (Goswami & Brown, 1990), associative inference (Preston &

Eichenbaum, 2013), induction (Gelman & Markman, 1986), and self-derivation (Bauer & San Souci, 2010; Bauer & Varga, 2017).

Self-derivation of new factual knowledge is of particular interest for several reasons. Unlike other productive process paradigms such as associative inference (Schlicting, Zeithamova, & Preston, 2014; Zeithamova & Preston, 2017), the self-derivation paradigm operates over meaningful factual content. This makes it an especially viable candidate model for knowledge accumulation. For example, in one episode of learning you may learn the "stem" fact *Duchamp's most influential work is called Fountain*. In a later episode, you learn the related stem fact *A popular sculpture made from a urinal is called Fountain*. Integration across these separate episodes allows for the derivation of the new factual knowledge *Duchamp's most influential work consisted of a urinal*. Its products are quickly incorporated into the knowledge base (Bauer & Jackson, 2015). Both children (Dugan & Bauer, 2020; Varga & Bauer, 2013) and adults (Varga & Bauer, 2017a) retain information learned through self-derivation for at least one week. Self-derivation is also a flexible process that allows for knowledge generation in a multitude of topics and across different presentation mediums. To date, self-derivation has been observed in the laboratory (Varga & Bauer, 2017a; 2017b), the classroom (Esposito & Bauer, 2017, 2019), and museum exhibits (Cronin-Golomb & Bauer, 2020), and it is clear that self-derivation operates over a variety of content including topics that are often reserved for learning from "informal" sources outside of the classroom (e.g., prescription medications; Dugan & Bauer, 2020). Likely due to the factual content over which participants self-derive, self-derivation performance is unsurprisingly related to academic achievement in both elementary and college students (Varga, Esposito, & Bauer, 2020).

Multiple studies have revealed substantial individual variability in self-derivation performance, warranting further investigation into the factors that contribute to these individual differences (Varga & Bauer, 2017a, Varga, Esposito, & Bauer, 2020). As one example, Varga and Bauer (2017a) found that self-derivation performance in their sample of academically high-achieving undergraduate students ranged from 3% to 93% correct, with 50% correct on average. These participants also completed tasks assessing verbal comprehension, working memory, short term memory span, and visual-auditory learning. Both working memory and verbal comprehension were found to be unique contributors to the observed variability in self-derivation of new factual knowledge. More specifically, Varga and Bauer (2017a) found that verbal comprehension and working memory were the only significant predictors of self-derivation performance from amongst a battery of cognitive measures. This work emphasizes the presence of individual differences and suggests that the variability in self-derivation performance relates to other cognitive measures in meaningful ways). These studies highlight the need to further investigate the factors underlying these differences.

One potential source of individual differences in self-derivation is metacognition. Metacognition is cognition about one's cognition or "thinking about thinking" (Flavell, 1979). Metacognition may provide one the awareness to guide learning in service of successful self-derivation of new knowledge. As outlined in the influential model by Nelson and Narens (1990), there are two main metacognitive processes: monitoring and control. Metacognitive monitoring involves what we know about our cognition. This knowledge comes from observing and making judgments about how well an item has been learned, for example. It includes the feelings experienced when acquiring information or attempting to retrieve it such as lacking confidence in one's performance or feeling certain of a particular response (Flavell, 1979). The second

process, metacognitive control, refers to the many steps we take to modify our cognition including looking for strategies that will improve learning outcomes, initiation of a strategy to test its efficacy, and eventual termination of that strategy when one's learning goal has been realized or when it is clear that attempt at regulation was unsuccessful. Monitoring is evaluated through elicited metacognitive judgments, whereas attempts to understand control have included measures of study duration, strategy selection, and item selection. A critical aspect of the Nelson and Narens (1990) model of metacognition is the interaction between monitoring and control processes: each can inform and guide the other, thereby forming a feedback loop of information about cognition. Monitoring and control can influence learning processes at nearly any point during acquisition, retention, or retrieval.

To date there has been no work directly testing relations between productive processes and metacognition. There is suggestive evidence of such relations, however. Adults who described having insight into a novel transitive inference task performed significantly better than those who were not aware of the task structure and demands (Smith & Squire, 2005). Evidence suggestive of a similar pattern of greater awareness into task structure relating to higher performance is also present in two studies of adults' self-derivation (Varga & Bauer, 2017; Menkes, Esposito, & Bauer, 2018). Following test for self-derivation, Varga and Bauer (2017a) consulted adults' knowledge of relatedness between pairs of facts. They found a moderately strong relation between knowledge of relatedness and successful self-derivation, $r = .44$. Furthermore, participants who were unsuccessful on a subset of trials but reported being aware of the opportunity to integrate exhibited longer response times specifically on unsuccessful trials. This suggests participants engaged in persistent search of their memories for a correct response, perhaps because they had a sense of putting facts together across the session (Varga & Bauer,

2017a). Additional evidence of task awareness impacting performance comes from adults who engaged in multiple blocks presenting the opportunity to self-derive (i.e., encode—test; encode—test; etc.) (Menkes et al., 2018). Performance improved from the first block to subsequent blocks, despite participants never receiving instructions on how to perform successfully (Menkes, et al., 2018). Taken together, these data highlight that not only is there individual variability in self-derivation performance but that this variability is likely related, at least in part, to differential metacognitive monitoring and control as self-derivation unfolds.

As metacognition involves both monitoring and control processes, metacognition may be expected to influence self-derivation at any step in the process of self-derivation. Self-derivation unfolds over time, providing ample opportunity for metacognitive monitoring to influence performance likely in service of successful knowledge extension. The ERISS model (Bauer & Varga, 2017), a process model of self-derivation, highlights the various opportunities that metacognition may impact and contribute to individual variability in performance. According to the ERISS model, first an initial fact *Duchamp's most influential work is called Fountain* is Encoded. This fact is then Reactivated upon encoding of a second, related fact *A popular sculpture made from a urinal is called Fountain*. These two facts are then Integrated in memory with each other based on the shared information *Fountain*, as well as with any prior knowledge of dolphins. Upon a demand – *What did Duchamp's most influential work consist of?* – relevant information necessary to meet the demand is Selected. The new knowledge that pods talk by clicking and squeaking is then Self-derived. This model highlights a critical observation about integration and self-derivation, namely, that the many subprocesses involved unfold over time. There is some evidence that variability in successful engagement of the steps contributes to both age-related change and individual variability in successful self-derivation (see Bauer, Cronin-

Golomb, Porter, Jaganjac, & Miller, 2020, for discussion). Metacognition may influence self-derivation by contributing to the variability in one or more of the subprocesses outlined by ERISS, but because metacognition has yet to be studied in the context of self-derivation through memory integration, we focus here on the impact of metacognitive monitoring upon encoding of new information.

Metacognition is not only a resource for learners to utilize as they study or acquire new knowledge; it is also a tool researchers can employ to better understand cognition. One particularly good measure of metacognition is the judgment of learning. Judgments of learning (JOLs) are prospective judgments of how well one believes she or he has learned either individual items or a set of items. Judgments of learning can be made immediately after studying items or following a delay. For example, in a metamemory study with paired associates a participant will see a pair of words, *apple-blanket*, on a computer screen. Once the pair disappears, participants are then asked, "What are the chances that you will recall the second word when presented with the first word?" (Koriat, 1997). Participants indicate their responses as percentages between 0% and 100% certainty. Accuracy of item-by-item JOLs improves with repeated studying. That is, as participants study more, their predictions of their subsequent memory improve (Koriat, 1997). Koriat (1997) posits that repeated studying improves JOL accuracy because with continued study participants begin relying on internal cues about memory for the stimuli, suggesting that they attend to how easy or hard it is to remember those items. If participants attend to their mnemonic cues about stimuli, then in the context of self-derivation they may attend to the relatedness of the separate stem facts or the integrative nature of the task overall.

Accuracy of metacognitive judgments like JOLs is typically quantified by gamma correlations, a measure calculated by correlating the concordance (or discordance) of a metacognitive judgment with its related item. Metacognitive accuracy is a logical predictor of academic achievement because unless learners ignore their metacognitive knowledge, judgments can guide learners toward more accurate performance. In line with this Nietfeld, Cao, and Osborne (2005) found that undergraduate students' metacognitive accuracy was stable over the course of a semester and related to test performance despite changes in test difficulty. These findings highlight the utility of JOLs as windows into participants' learning. Judgments of learning are the most widely-used judgments (Koriat, 2007) and a viable candidate for better understanding the process of self-derivation.

In the present research, as a first step to understanding the influence metacognition may have on self-derivation, we investigated whether making a metacognitive judgment at encoding is related to subsequent self-derivation. Encoding of stem facts is the first opportunity for metacognition to directly impact a particular item in learning. Differential encoding impacts subsequent memory of both item and associative memories (Becker, Kalpouzos, Persson, Laukka, & Brehmer, 2017). From previous research on self-derivation through integration, it is already known that individuals who recall both stem facts necessary to self-derive new knowledge are more likely to be successful at self-derivation, but it is also clear that memory for both stem facts is not sufficient for self-derivation (Bauer et al., 2012; Bauer & San Souci, 2010). JOLs made during stem-fact encoding will provide evidence necessary to inform our understanding of metacognition's influence on self-derivation performance.

In the present research, participants encoded 44 pairs of related stem facts. To test the impact of metacognitive judgments at encoding on self-derivation, participants completed trials

in each of four conditions: judgment of learning for the first stem fact of a pair only, judgment of learning for the second stem fact of a pair only, judgment of learning for both stem facts in a pair, and no judgment for either stem fact in a pair. We then compared self-derivation and stem-fact memory performance across these four conditions. We hypothesized that if making a metacognitive judgment influences self-derivation performance then participants would be more successful on trials when they had judged one or both of the stem facts necessary to self-derive compared to trials that did not involve judged stem facts. Similarly, we expected that if making a judgment impacts memory for the stem facts, then participants would remember more of the judged stems compared to the stems they did not judge.

In summary, the primary motivation of the present research is to elucidate whether explicit metacognitive monitoring impacts later self-derivation. Such monitoring may contribute to the individual variability exhibited by adults as they engaged in the process of self-derivation of new factual knowledge through integration of separate yet related episodes of new learning. To achieve this aim, we presented participants with novel pairs of facts that can be integrated and then stored in memory. As participants engaged in the task, they provided judgments of learning. We then assessed self-derivation and memory for the individual stem facts.

Method

Participants

One hundred and one undergraduate students ($M = 19.26$ years, $SD = .94$; 70 females) participated in the study. An additional 14 adults participated but were excluded for having participated in a related study. Participants were recruited from introductory psychology courses at a mid-sized private research university. They received course credit for participation. Written informed consent was obtained prior to the start of the study session. Based on a self-report

demographics questionnaire, the sample was 1% American Indian or Alaskan Native, 23% Asian, 15% Black or African American, 58% White or Caucasian, and 3% multi-racial. Approximately 10% of the sample self-identified as being Hispanic or Latino. All participants were self-identified native speakers of English. The protocol and procedures for this experiment were approved by the university Institutional Review Board.

Materials

A corpus of 44 pairs of integrable facts (i.e., 88 "stem" facts) was used. A subset of 30 fact pairs was used in prior research (Varga & Bauer, 2017a). The remaining fact pairs were pilot tested and validated as requiring integration for successful self-derivation of new knowledge: participants' self-derivation performance was significantly higher when given both members of a stimulus pair compared to when they received only one stem fact from a pair, but not both. Facts were from a variety of domains including history, science, literature, and culture. Facts ranged from 3 to 13 words in length and were an average of 7.96 ($SD = 1.97$) words long. Facts were an average Flesch Kincaid reading level of 8.87 ($SD = 4.03$). An example pair includes the stem facts *The most intelligent invertebrates are called cephalopods* (Stem 1) and *Cuttlefish are considered to be cephalopods* (Stem 2). These can be integrated to self-derive the integration fact *Compared to other invertebrates cuttlefish are more intelligent.*

Design and Procedures

Participants were tested individually or in groups ranging from 2 to 10 participants (*M*=6) to maximize testing efficiency. All participants completed each task independently. In Phase 1 of the session, participants encoded 44 pairs of related stem facts via Qualtrics surveys. Each stem fact was presented as a single sentence on the screen. Members of a pair were separated by an intervening 2-10 facts ($M = 5.38$, $SD = 1.81$). Participants provided metacognitive judgments of

learning (JOLs) for half of all stem facts across 4 within-subjects conditions. Judgments were provided immediately after reading the stem fact but while the sentence was still visible on the screen. Specifically, participants responded to the question "How well do you think you will remember this fact later on?" on a scale of 0 to 100. They were first shown an example fact "George Washington was the first U.S. president" and then shown the judgment scale. Participants were instructed that they were free to choose anywhere on the scale with 0 reflecting they were not sure at all and 100 reflecting complete confidence they would remember the stem facts at test. As reflected in Table 1, each participant provided a total of 44 JOLs: 11 judgments for Stem 1 facts (Stem 1 Only condition) and 11 judgments for Stem 2 facts (Stem 2 Only condition). The remaining 22 judgments were given for both members of 11 integrable pairs (Both Stems condition). Which facts prompted JOLs was counterbalanced across the sample. Pairs across the 4 judgment conditions were pseudorandomly distributed across the encoding task to lessen the likelihood that participants would discern a pattern of what facts prompted judgments. Following the encoding task, participants completed unrelated buffer tasks as a group for a duration of approximately 25 minutes.

In Phase 2, participants individually completed new Qualtrics surveys comprised of fill-in-the-blank sentences testing for self-derivation of new knowledge. Participants' responses were scored online by the surveys. Participants then completed fill-in-the-blank sentences testing memory for the individual stem facts. Following open-ended testing, participants answered forced-choice questions for all integration questions answered incorrectly in open-ended format and then all stem fact questions answered incorrectly in open-ended format. Each forced-choice question had 4 alternatives, one of which was correct. The alternatives included at least one choice encoded in another fact pair and 1-2 plausible but novel distracters the participants had

not experienced in the session. Familiar options from other fact pairs were used approximately equally often across all distracter choices. We followed the same criteria for self-derivation and stem fact questions. For example, for the self-derivation question *Compared to other invertebrates, cuttlefish are more _____*, the answer choices were a) intelligent, b) poisonous, c) flexible, and d) endangered. Order of answer choices was counterbalanced across the sample of participants.

In the third and final phase of the experiment, participants completed demographic and debriefing questionnaires. In the debriefing questionnaire, participants were asked to state which facts they knew before participating in the study and to describe any strategies they used over the course of the encoding and test portions of the experiment. According to self-report, 97% of participants knew only a few or some of the facts before participating in the study. In response to the question, " Did you have the impression that there was some kind of logical rule or operation in the study? If so, please explain briefly" 49% of participants described noticing a relation or putting information together over the course of the study. Ten participants specifically described putting two separate facts together to learn new information.

Results

Self-derivation Performance

Participants self-derived an average of 18.73 (42.57%) integration facts in open-ended testing out of 44 trials ($SD = 8.24$). We calculated a total-score by adding the correct open-ended performance to unique correct forced-choice performance. Participants scored an average of 34.94 (79.41%) self-derivation trials correct in total performance on 44 trials ($SD = 5.86$).

There was a significant main effect of judgment of learning condition on open-ended self-derivation performance, $F(3, 300) = 5.602$, $p < .001$, $\eta_p^2 = 0.053$. As suggested by

inspection of Figure 2, Bonferroni corrected pairwise comparisons revealed that making a judgment of learning for one or both of the facts in a stem-fact pair resulted in significantly higher self-derivation performance than when judgments were not made at encoding. There were no other differences in open-ended self-derivation performance by JOL condition.

Total self-derivation performance also differed as a main effect of JOL condition $F(3,300) = 3.502$, $p = .016$, $\eta_p^2 = 0.034$. Bonferroni corrected pairwise comparisons showed that making a judgment of learning for either both facts in a stem fact pair or for Stem 2 resulted in significantly higher total self-derivation performance than when not making a judgment. There were no other significant differences in total self-derivation performance.

Stem Fact Memory

Participants remembered an average of 52.29 stem facts (59.45%) in open-ended testing ($SD = 14.25$) and 75.21 stem facts (85.48%) in total performance ($SD = 9.22$).

Participants remembered significantly more stem facts for which they made JOLs during encoding than those stem facts which they did not; $t(100) = 5.239$, $p < .001$, $.d = 0.52$. Stem fact performance is shown in Figure 3.

When open-ended stem fact performance was analyzed across all JOL conditions there was a main effect $F(2.77, 276.967) = 9.26$, $p < .001$, $\eta_p^2 = 0.085$, such that proportion correct stem fact memory was higher when a judgment of learning was made than when it was not. Bonferroni corrected pairwise comparisons revealed that proportion correct stem fact memory for all three conditions involving JOLs was significantly higher than the no-judgment condition. The three JOL conditions did not differ amongst themselves.

Stem fact performance was also analyzed for an effect of temporal order. Paired sample t-test revealed no difference in participants' memory for the first encoded stem fact in a pair ($M = .$

59, *SD* = .17) relative to the second encoded stem fact in a pair (*M* = .60, *SD* = .17). There was no difference; *t*(100) = -.182, *p* = .856, *d* = -0.02.

Judgments of Learning and Metacognitive Accuracy

On a scale of 0 to 100, participants gave an average JOL of 53.08 (*SD* = 12.26) reflecting moderate confidence in future memory performance. The majority of participants utilized the full range of the scale; 4 of 101 participants (3.9%) utilized only the upper half of the scale.

For pairs wherein participants made JOLs for both stem facts, we examined whether the second JOL made for a given stem fact pair was higher than the first. We hypothesized that if participants monitor learning at the level of the individual fact, then participants may provide higher JOLs for the second stem because they recognize they have encountered both facts. We found that JOLs made for the second encoded stem fact were higher than those made for the first encoded stem fact , *t*(100) =3.754, *p* < .001, *d* = .37.

We calculated the accuracy of participants' JOLs for the individual stem facts by conducting a gamma correlation between JOL and subsequent stem fact memory. Participants' average gamma for JOLs across all three judgment conditions was *G* = 0.51 (*SD* = .22). This indicates that participants accurately predicted their memory for approximately half of all judged stem facts. When comparing across the JOL conditions there was a main effect of condition; $F(2,182) = 7.710$, *p* = .001, $\eta_p^2 = 0.078$. Participants exhibited significantly greater accuracy for the Stem 2 only condition (*M* = .59, *SD* = .32) compared to the Stem 1 only condition (*M* = .40, *SD* = .45). The Both Stem condition was intermediate between the single stem conditions (*M* =.50, *SD* = .27) and was not significantly different from either. Because stem facts were counterbalanced for both temporal order and which fact contained the key words necessary to

answer the self-derivation question (Stem 1/Stem 2), it is unclear what accounts for this difference.

We also utilized JOLs for stem facts to analyze accuracy of JOLs predicting subsequent self-derivation performance. For JOLs in the Both Stem condition, we calculated the mean JOL between the stems in a pair. Overall, participants' displayed an average gamma correlation of $G = .35$ for how accurate their JOLs given after stem facts were in relation to subsequent open-ended self-derivation performance (OE). There was no difference in the gammas depending on whether the participant made a JOL for Stem 1 ($M = .38$, $SD = .37$), Stem 2 ($M = .30$, $SD = .42$), or both stem facts ($M = .40$, $SD = .42$); $F(2, 182) = 1.934$, $p = .148$, $\eta_p^2 = 0.021$.

Discussion

The present research was the first direct examination of the relation between self-derivation of new factual knowledge and metacognition in adults. The aim was to determine whether metacognitive monitoring during encoding of separate yet related stem facts impacts subsequent self-derivation. Participants made judgments of learning at encoding for none, one, or both of the stem facts in a related pair. They later answered open-ended questions assessing their self-derivation of new knowledge and memory for the individual stem facts. Based on prior research, we hypothesized that if participants engage in metacognitive monitoring then self-derivation performance would be positively influenced.

Adults in the present research were quite successful at self-deriving new knowledge and remembering the individual stem facts. Adults successfully self-derived on 43% of open-ended trials and 79% of total trials. They also remembered 59% of stem facts in open-ended testing and 85% in total performance. This high performance in both self-derivation and stem fact performance is due in part to the addition of metacognitive judgments at encoding. Self-

derivation and stem fact performance were both higher when participants made judgments of learning about both stem facts within a pair. Thus, as hypothesized, metacognitive monitoring positively influenced self-derivation performance.

As has been observed in previous research (Varga & Bauer, 2017a; Varga, Esposito, & Bauer, 2020), adults exhibited successful self-derivation performance ranging from 6% to 84% correct. Despite the increased cognitive demand of 44 self-derivation trials compared to the previous maximum of 30 trials in prior research (Varga & Bauer, 2017a), we observed a similar pattern of results. This highlights the pervasiveness of self-derivation as a means of knowledge accumulation and that variability in self-derivation is observed across a range of conditions. Moreover, it is noteworthy that adults in the present research varied nearly as much as those who participated in prior work (Varga & Bauer, 2017a), even in the face of higher performance as a result of making Judgments of Learning.

Few studies have investigated the impact of making a metacognitive judgment on subsequent memory (Dougherty, Robey, & Buttaccio, 2018; Dougherty, Scheck, Nelson, & Narens, 2005) but much of the evidence from these experiments suggests that metacognitive judgments do not improve subsequent memory. Memory is improved more by retrieval practice rather than by metacognitive judgments. There are several key points to consider regarding how these studies align with the present work. First, in Dougherty and colleagues (2005, 2018), participants made judgments of learning before the final test but after already having the opportunity to retrieve the target information. In the present work, participants made JOLs during initial encoding of the stem facts, so each judgment was intended to monitor present learning and predict future performance. Secondly, in contrast to both studies by Dougherty and colleagues (2005; 2018), the present research is a within-subjects comparison making the

difference between judgment and no-judgment conditions more striking. It is possible that participants could have carried the practice of making a metacognitive judgment throughout all completed trials, but the consistent finding that both self-derivation and stem fact memory performance were higher when participants made directly related judgments suggests that only elicited JOLs impacted performance. Finally, the nature of the materials presented in Dougherty et al. (2018) and Dougherty et al. (2005) is crucially different from the information with which participants interacted in the present research. That is, Dougherty and colleagues (2005; 2018) studied metacognition for word pairs whereas we investigated metacognition for pairs of related facts. Retrieval practice may be necessary to improve memory in the paradigm used by Dougherty and colleagues (2005; 2018) for the target words in a paired-associate paradigm because the word pairs are otherwise arbitrary to participants. In contrast, stem and integration facts learned through the course of the present experiment are intended to be novel but ecologically relevant facts that one could encounter over the course of their educational or everyday experiences.

Familiarity of encoded information may impact one's judgments such that higher familiarity contributes to higher judgments of learning (Schwartz & Metcalfe, 1992; Son & Metcalfe, 2005). This is important to consider for the present work because the very nature of integrable pairs of facts relies on the shared information, the link between facts in a pair. We hypothesized that JOLs made to the second stem fact will likely be higher for those who notice, on some level, that there is a link between members of a pair. Alternatively, it is possible that perceived ease of comprehension at the second stem fact, based on this familiarity, actually poses a threat to successful self-derivation (Varga & Bauer, 2017b). In previous research of adults' self-derivation, Varga and Bauer (2017b) recorded participants' event-related potentials as

participants encoded pairs of integrable facts. On trials when participants subsequently self-derived new knowledge, the second stem in a pair was marked by a significantly more positive amplitude than was observed for Stem 1. When self-derivation was not successful, there was no difference in the amplitude of the ERPs for Stem 1 and Stem 2. Varga and Bauer (2017) suggest that Stem 2 is recognized as new information prior to successful self-derivation; in contrast Stem 2 is treated as familiar or old on subsequently unsuccessful trials. Importantly, these participants did not make explicit judgments about their learning. The present finding that adults did not give higher JOLs for the second stem in a pair complements the ERP data collected by Varga and Bauer (2017b) as it suggests participants treated Stem 2 as new rather than dismissing it as familiar. Taken together, these data provide support that metacognition at encoding of related stem facts positively influences subsequent self-derivation performance.

The impact of metacognition on self-derivation has never before been directly examined. Though this study is but the first investigation toward understanding how metacognition contributes to variability in self-derivation and the work was not designed to discover a strategy for improved performance, the results are promising. What may be considered a simple intervention has a meaningful impact on participants' learning; making one or more judgments while learning the separate stem facts contributed to significantly higher self-derivation performance. Additionally, participants were better able to remember the stem facts when they made a judgment about them. This application will require further work to really test.

Limitations and Future Directions

The present results suggest that metacognitive monitoring impacts successful self-derivation. It should be noted that the within-subjects design of the present work lends support to the interpretation of the significant difference between judged and non-judged stem facts and

self-derivation trials. However, there remains the possibility that participants' monitoring during encoding of the stem facts carried over to facts that were not paired with an elicited judgment. A between-subjects design reflecting a similar pattern of results as the present research would provide support necessary to further disentangle the role that metacognition plays in self-derivation.

Future work will test for replication and extension of the present findings by investigating the relation between the individual judgments of learning provided at encoding and the relative success of subsequent self-derivation. As is evident by the ERISS model (Bauer & Varga, 2017), there are multiple opportunities to prompt for metacognitive judgments and ultimately determine both the number and timing of metacognitive judgments that contributes to highest self-derivation performance. Future investigations will also test the relation between metacognition and retention of self-derived knowledge.

Conclusion

In summary, the present study employed Judgments of Learning to investigate the relation between metacognitive monitoring during encoding of related stem facts and subsequent self-derivation performance. Results indicated that making JOLs about stem facts increased both self-derivation and stem fact memory compared to stem fact pairs that were not judged at encoding. This contributes both to our understanding of metacognition in the context of self-derivation as well as the influence of eliciting metacognitive judgments during memory-related tasks more generally. These findings suggest that metacognition does influence self-derivation performance and that future work should further elucidate how metacognition may contribute to the individual variability observed in adults' self-derivation.

References

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. doi:10.1037/a0037405

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

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., & Varga, N. L. (2017). Similarity and deviation in event segmentation and memory integration: Commentary on Richmond, Gold, & Zacks. *Journal of Applied Research in Memory and Cognition*, *6*(2), 124-128.

Becker, N., Kalpouzos, G., Persson, J., Laukka, E. J., & Brehmer, Y. (2017). Differential effects of encoding instructions on brain activity patterns of item and associative memory. *Journal of Cognitive Neuroscience*, *29*(3), 545-559.

Dougherty, M. R., Robey, A. M., & Buttaccio, D. (2018). Do metacognitive judgments alter memory performance beyond the benefits of retrieval practice? A comment on and replication attempt of Dougherty, Scheck, Nelson, and Narens (2005). *Memory & Cognition*, *46*(4), 558-565.

Dougherty, M. R., Scheck, P., Nelson, T. O., & Narens, L. (2005). Using the past to predict the future. *Memory & Cognition*, *33*(6), 1096-1115.

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

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*(10), 906.

Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition, 23*(3), 183-209.

Goswami, U. (2011). Inductive and deductive reasoning. In U. Goswami (Ed.), Childhood Cognitive Development (pp. 399–419). Oxford, UK: Wiley–Blackwell. doi.org/10.1016/j.tics.2010.10.001

Goswami, U., & Brown, A. L. (1990). Higher-order structure and relational reasoning: contrasting analogical and thematic relations. *Cognition, 36*(3), 207–226.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*(4), 349.

Koriat, A. (2007). *Metacognition and consciousness.* In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge handbook of consciousness* (p. 289–325). Cambridge University Press. https://doi.org/10.1017/CBO9780511816789.012

Menkes, M. W., Esposito, A. G., & Bauer, P. J. (2018). Self-Generation of Novel Information through Integration of Cross-Language Episodes. *Manuscript in preparation.*

Nietfeld, J. L., Cao, L., & Osborne, J. W. (2005). Metacognitive monitoring accuracy and student performance in the postsecondary classroom. *The Journal of Experimental Educational*, 7-28.

Nelson, T.O. & Narens, L. (1990). Metamemory: A theoretical framework and some new

findings. In G.H. Bower (Ed). The Psychology of Learning and Motivation, 26, 125-173.

New York: Academic Press

Preston, A. R., & Eichenbaum, H. (2013). Interplay of hippocampus and prefrontal cortex in

memory. *Current Biology*, *23*(17), R764-R773.

Schlichting, M. L., Zeithamova, D., & Preston, A. R. (2014). CA1 subfield contributions to

memory integration and inference. *Hippocampus*, *24*(10), 1248-1260.

Schwartz, B. L. (1994). Sources of information in metamemory: Judgments of learning and

feelings of knowing. *Psychonomic Bulletin & Review*, *1*(3), 357-375.

Schwartz, B. L., & Metcalfe, J. (1992). Cue familiarity but not target retrievability enhances

feeling-of-knowing judgments. *Journal of Experimental Psychology: Learning, Memory,*

and Cognition, *18*(5), 1074.

Smith, C., & Squire, L. R. (2005). Declarative memory, awareness, and transitive inference. *The*

Journal of Neuroscience, *25*(44), 10138-10146.

Son, L. K., & Metcalfe, J. (2005). Judgments of learning: Evidence for a two-stage

process. *Memory & Cognition*, *33*(6), 1116-1129.

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. (2020). Cognitive correlates of memory integration

      across development: Explaining variability in an educationally relevant

      phenomenon. *Journal of Experimental Psychology: General*, *148*(4), 739.

Zeithamova, D., Dominick, A. L., & Preston, A. R. (2012a). Hippocampal and ventral medial

      prefrontal activation during retrieval-mediated learning supports novel inference. *Neuron,*

      75(1)*,* 168-179.

Table 1

Judgments of Learning made per Condition

Condition	Number of Fact Pairs	Number of Judgments Made
Both Stems JOLs	11	22
Stem 1 Only	11	11
Stem 2 Only	11	11
No JOL	11	0
Total	44 Pairs	44 JOLs

Table 2

Mean Open-ended and Total Self-derivation Performance by Condition

Condition	Self-derivation Mean (SD)	Total Self-derivation Mean (SD)
Both JOLs	5.10 (2.43)	8.95 (1.67)
Stem 1 JOL	4.66 (2.50)	8.79 (1.87)
Stem 2 JOL	4.70 (2.33)	8.79 (1.81)
No JOL	4.27 (2.37)	8.41 (1.92)
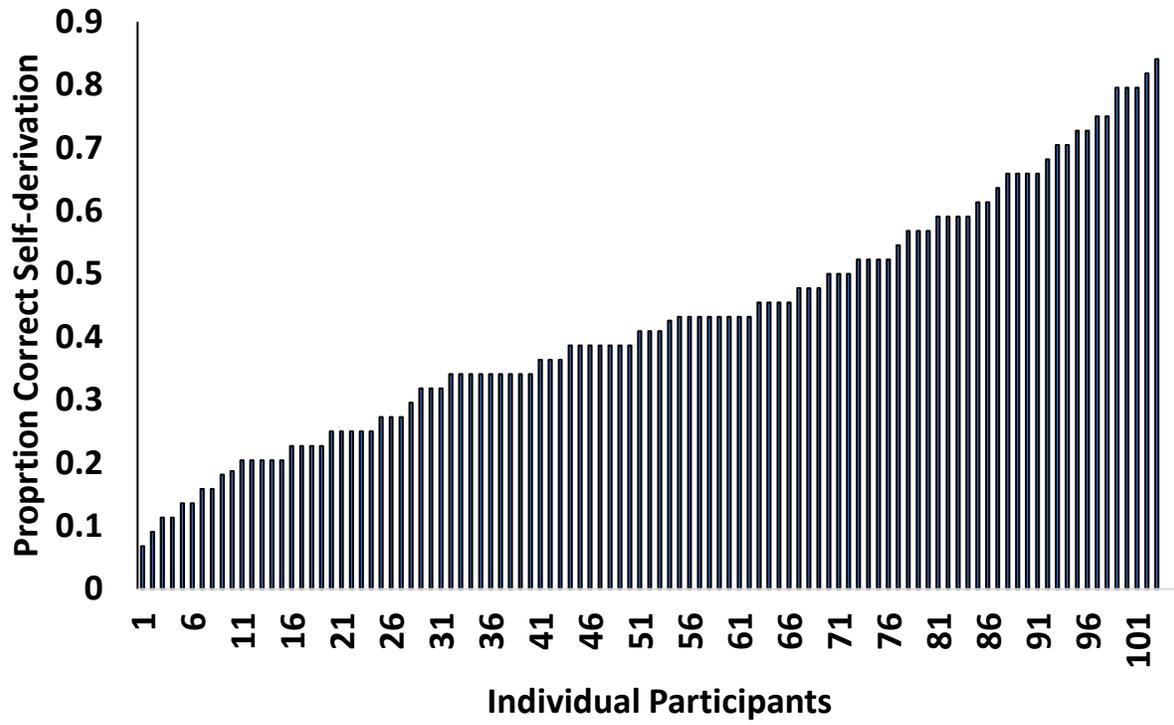
Figure 1. Range of Correct Self-derivation Performance across all Participants
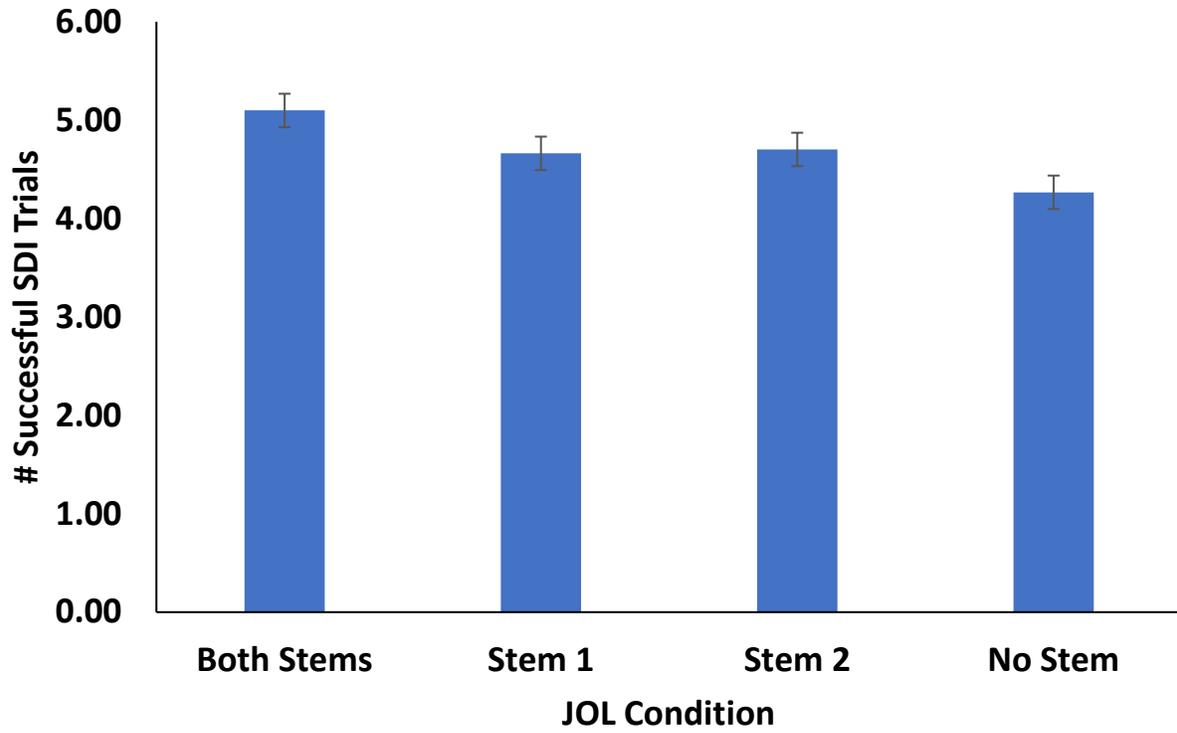
Figure 2. Self-derivation Performance by Judgment of Learning Condition
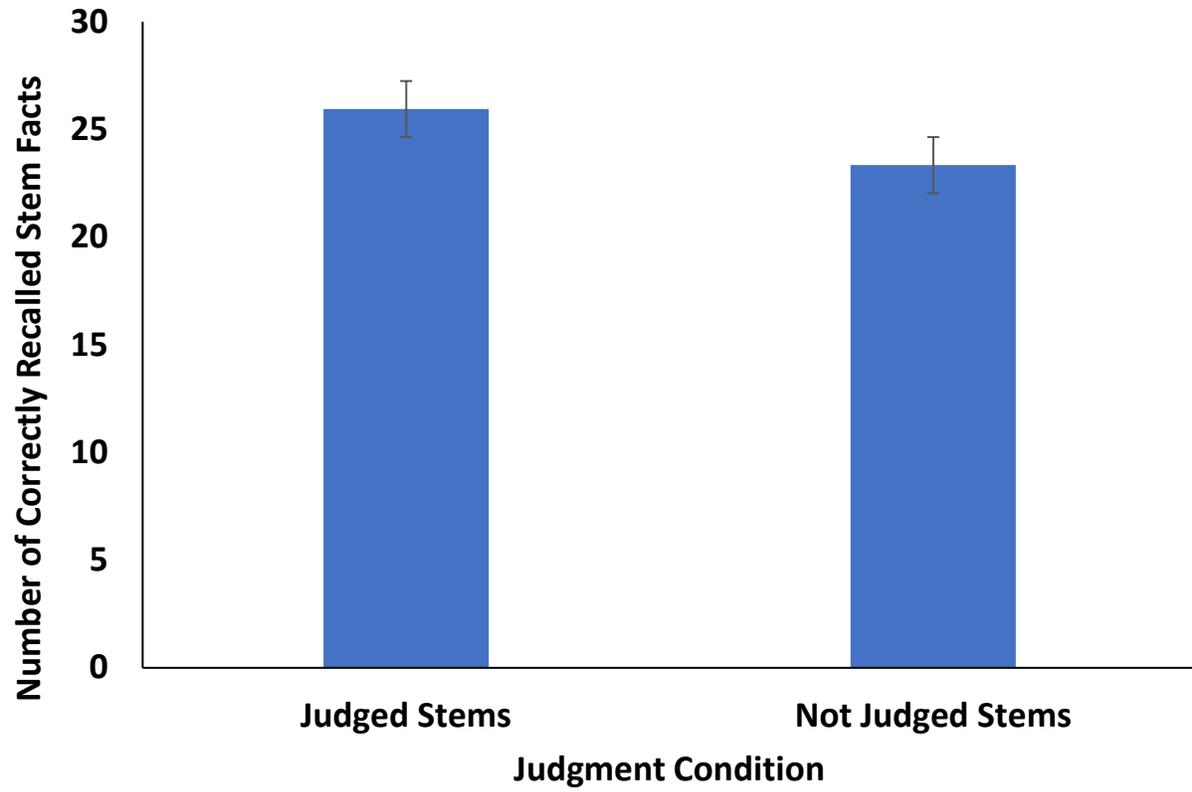
Figure 3. Stem fact Memory by Judgment Condition

Appendix

Additional Analyses

We also analyzed the relation between time taken to make a Judgment of Learning and the magnitude of the judgment. The average correlation between latency and JOL was $r = -.05$ ($SD = .19$). Latency is sometimes used as an indirect measure of certainty as longer latencies suggest longer memory search or vacillation between answer choices.

Study 2

Putting the pieces together: A cross-sectional investigation of children's metacognition about

self-derivation

Jessica A. Dugan & Patricia J. Bauer

Emory University

Putting the pieces together: A cross-sectional investigation of children's metacognition about self-derivation

One of the central goals of development is to build a knowledge base. The learning that is necessary for accumulating knowledge occurs via different means and opportunities including, but not limited to, watching educational television shows, exploring museum exhibits, reading books, or engaging in classroom lessons. Importantly, any single instance of learning is not sufficient for the accumulation of knowledge. To build a knowledge base, one must integrate information over time. Extending beyond what is directly provided in a learning episode allows knowledge accumulation to be efficient. Productive processes allow for such extension and contribute to the accumulation of knowledge; these processes include deduction, analogy, and the focus of the present work – self-derivation. Self-derivation is the productive process wherein one can derive new knowledge via integration of separate yet related episodes in memory (Bauer & San Souci, 2010; Bauer & Varga, 2017). Importantly, knowledge accumulation is not supported by productive processes alone. Learning is also supported by metacognition, the knowledge and regulation of one's cognition (Flavell, 1979; Hacker, Dunlosky, & Graesser, 2009; Nelson & Narens, 1990, Metcalfe & Finn, 2008). Though there are hints that adults' metacognitive awareness impacts performance on productive processes such as transitive inference (Smith & Squire, 2005), how metacognition supports self-derivation of new knowledge specifically and whether the potential relation between these two cognitive abilities may vary across development are currently unknown. To address these questions, in the present research, we investigated metacognition about learning through self-derivation in children grades 3 through 5 and examined the potential relation between these two ecologically relevant processes.

Productive processes contribute significantly to knowledge base development. They include, but are not limited to, analogy, induction, transitive inference, associative inference (see Goswami, 2011, for a review) and self-derivation. Self-derivation is a process that unfolds over multiple subprocesses and occurs through the integration of separate yet related episodes of new learning (Bauer & Varga, 2017). To illustrate, children were presented with the fact "dolphins talk by clicking and squeaking" (Stem 1) in one passage. From a later passage, children learned the stem fact "dolphins live in groups called *pods*" (Stem 2). To gain the knowledge needed to answer the question "How does a pod talk?" children must have integrated the separate yet related stem facts. Crucially, self-derivation is an especially salient target for investigation for at least 5 reasons. First, the materials over which self-derivation operates are novel but true factual information. Participants engaged in self-derivation learn pairs of facts about topics such as geography, history, chemistry, literature, and music (Bauer & Jackson, 2015; Varga & Bauer, 2017a, 2017b). Second, the factual knowledge gained via self-derivation through integration is not only ecologically relevant; it is also rapidly incorporated into the knowledge base (Bauer & Jackson, 2015). Third, self-derivation is also significantly related to academic achievement in both adults and children (Varga, Esposito, & Bauer, 2020). Fourth, self-derived knowledge endures delays of at least one week (Dugan & Bauer, 2020; Varga & Bauer, 2017a; Varga, Stewart, & Bauer, 2016). Finally, the new knowledge that arises from self-derivation is internally generated in the mind of the child or adult, thus making metacognition uniquely suited to provide insight into what the learner understands of this process. Although these qualities may be true for other forms of productive processes (e.g. associative inference), that has yet to be tested.

Studies spanning childhood reflect that self-derivation develops substantially across the elementary school years (Bauer & Larkina, 2017; Bauer & San Souci, 2010; Varga, Stewart, &

Bauer, 2016). In a text passage format of the self-derivation paradigm, children were read pairs of related passages (Bauer & Larkina, 2017). Each passage featured a true but novel fact ("stem fact"). In this paradigm, 4-year-olds successfully self-derived new knowledge on 13% of trials whereas 8-year-olds were successful on 75% of trials, a very robust increase compared to their younger counterparts. Importantly, when children are provided with only one of the stem facts from a pair, they do not go on to self-derive new knowledge (Bauer & Larkina, 2017; Bauer & San Souci, 2010).When children in grades K through 3 engage in the story-passage paradigm in their home classrooms, lower levels of open-ended self-derivation are observed, likely due in part to the distractions that are inherent in a classroom environment (e.g., classmates, technology, or an unfinished homework assignment). However, a similar pattern of age-related change in self-derivation emerges even in a classroom setting (Esposito & Bauer, 2017). Across development, self-derivation serves as an educationally relevant and ecologically valid mechanism of rapid knowledge accumulation in both the laboratory and in children's familiar classrooms.

In addition to age-related change in self-derivation performance, there is also individual variability in successfully deriving new knowledge. Especially because learners are themselves the authors of their own knowledge generated through self-derivation, a likely source of variability in the derivation and retention of new knowledge is metacognition. Metacognition is cognition about one's cognition or "thinking about thinking" (Flavell, 1979). Metacognition entails both the knowledge one has about her cognition and the regulation of cognition (Nelson & Narens, 1990; Schraw & Dennison, 1996). Knowledge about cognition comes from monitoring cognitive processes and their outcomes. For example, one may learn the separate stem facts of an integrable pair about dolphins. For each fact, he may try to predict how easily it

will be recalled later. Metacognitive knowledge can then be used to regulate cognition. Perhaps later upon learning Stem 2, the individual attempts to recall what else he knows about dolphins, including reactivating Stem 1. A person in this situation might then spend more time trying to rehearse and retrieve the stem facts to gain confidence in his future performance; such manipulation of the related stems may also result in the participant self-deriving new knowledge after integrating the facts in memory. Because self-derivation is a multi-step process that unfolds over time and results in a self-made product, there are many opportunities for metacognitive knowledge and regulation to influence self-derivation.

Though there are many opportunities at which metacognition may impact self-derivation, this has not yet been examined. Studies aimed at elucidating the relation between metacognition and other cognitive abilities often rely on measures of monitoring. Metacognitive monitoring is investigated and measured with metacognitive judgments (Dunlosky & Nelson, 1994; Hart, 1965; Koriat, 1997; Metcalfe & Finn, 2008). Participants may judge how well they believe they have learned a fact (Judgment of Learning), predict how likely they are to recognize an item on a later memory test (Feeling of Knowing), or rate their confidence after responding to a question (Confidence Judgment). These judgments can be made for a near-endless variety of stimulus types, though word pairs and lists are most frequently used. For example, in a study conducted by Finn and Metcalfe (2014), children in grades 3 and 5 learned definitions of words and then made judgments of learning as percentages between 0% and 100%. For metacognition to be efficacious, the judgments and actions born from monitoring and controlling cognition must be accurate. Relative accuracy of one's judgments is characterized by a gamma correlation (Nelson, 1984), measuring the proportion of judgments that are concordant with actual performance on a task. Positive gamma correlations approaching 1 reflect that a participant's judgments are

aligned with her actual performance whereas a negative gamma correlation would suggest a participant is more often judging her performance incorrectly.

Metacognition has not yet been investigated in the context of self-derivation but there are numerous ways it may impact one's successful self-derivation of new knowledge. Of critical importance for present purposes is that because the subprocesses unfold over time, there is ample opportunity for metacognition to influence performance. By combining the ERISS process model (Bauer & Varga, 2017) and work on metacognition, we can begin to investigate how metacognition is related to age-related variability in the process of gaining new knowledge via self-derivation of new knowledge through integration of separate episodes of learning. The ERISS model puts forth that self-derivation occurs over the course of 5 steps. Metacognition may impact self-derivation at any one (or more) of them. It might impact the first step in the process, Encoding (the "E" in ERISS), by directing attention to the shared relational information across integrable facts. Prior work makes clear that learning both stem facts is necessary but not sufficient for successful self-derivation (Bauer & San Souci, 2010; Bauer & Varga, 2017). Metacognitive monitoring may direct attention after encoding has occurred so that only relevant and related information is Reactivated (step 2) and subsequently Integrated (step 3). Metacognition may also strongly impact Selection (step 4) as it may control selection of necessary, relevant facts, from a corpus of knowledge, in service of deriving new understandings. Bauer and Larkina found 8-year-olds, but not younger children, self-derive new knowledge even in the presence of related distractor facts (2017). Finally, metacognitive monitoring and control may lead one to learn how to successfully Self-derive (step 5) through integration and provide the tool to probe one's knowledge of new self-derived understandings.

To understand the contributions metacognition may make to successful self-derivation, it is crucial to also understand how metacognition develops and functions outside of the context of self-derivation. Like self-derivation, metacognitive monitoring accuracy increases significantly between ages 4 and 10 years (Schneider, 1985). When making predictions about memory performance, young children in preschool and kindergarten tend to be overconfident in their abilities and do not perform as accurately as older children (Vise & Schneider, 2000). Accuracy of metacognitive monitoring, particularly predicting future performance, begins to improve beginning around age 7 years. Evidence from 7-year-olds suggests that children early in their educational careers can judge how well they will later remember an item and guide their studying toward the items that will be more difficult to remember (Tsalas, Paulus, & Sodian, 2015). By age 8 years most children judge accurately whether they will remember an item upon subsequent testing (Metcalfe & Finn, 2013) and accuracy is not significantly different from adults on tasks such as recalling details from an observed event and differentiating between answerable and unanswerable questions (Roebers, von der Linden, Howie, & Schneider, 2007). Longitudinal investigations of metacognitive development provide further support of the age-related increases in metacognitive monitoring accuracy (Roebers & Spiess, 2017; Roebers et al., 2019; Schneider & Lockl, 2002; Schneider & Loffler, 2016).

The Present Research

It is not yet known to what extent metacognition may be related to and support successful self-derivation through integration across separate episodes of learning. Metacognition may contribute to age-related variability in self-derivation given the observed age-related increases in metacognitive accuracy. It may impact self-derivation performance such that individuals with highly accurate metacognitive monitoring are more successful than those who do not use their

metacognition to guide cognition. The very nature of self-derivation relies on integrating information learned from separate episodes to generate new knowledge. It is reasonable to assume that an individual who is acutely aware of his learning and notices the information shared between two stem facts would experience greater success at deriving new knowledge. Conversely, one who monitors less accurately may not notice the integrable nature of the directly encoded stem facts. Importantly, self-derived facts are never presented directly to participants. This presents a unique opportunity to investigate children's insight into their own internal creation of new knowledge.

We aimed to investigate the relation between metacognition and self-derivation and to examine how this relation may differ with age, across grade level. To address these aims, we asked children to provide metacognitive judgments as they engaged in the processes of learning new information and then subsequently answering questions that could be addressed through reflection on that learning. We investigated this cross-sectionally in 3rd through 5th grade students in their classroom environments. The metacognitive aspects of the present research were added to an ongoing longitudinal study. To avoid influencing the results of the ongoing longitudinal study, all metacognitive judgments were elicited at test rather than at encoding of stem facts. Additionally, the goal of the present research is not to identify the particular ERISS step when metacognition may be at work.

Two metacognitive judgments particularly well-suited to the present investigation are feelings of knowing and confidence judgments. Feelings of knowing (FOK) rely on extrinsic properties of information and allow for understanding how well participants understand both their cognition and the contents of memory. Confidence judgments (CJs) elicited after learning, at test are generally thought to be based on the retrieval experience of testing (Nelson & Narens,

1990; Koriat, 2012), reflecting a veridical measure of participants' metacognitive monitoring. Due to the protracted yet seemingly simultaneous development of both self-derivation and metacognition, it is reasonable to assume that any relation between these processes may also change with age.

We conducted this work with students in grades 3-5 for three reasons: 1) children in this age range have previously shown evidence of successful self-derivation across multiple single-sentence trials (Bauer, Blue, Xu, & Esposito, 2017), 2) children in this age range can accurately monitor their learning under some conditions (Metcalfe & Finn, 2013; Roebers & Spiess, 2017; Roebers et al., 2019), and 3) these children were already participating in a broader longitudinal project including investigation of their self-derivation and academic achievement.

The present research serves as the first test for a relation between metacognition and self-derivation. We examined children's metacognitive judgments at test for self-derivation. To that end, because both metacognition and self-derivation undergo substantial change in the period between 3rd and 5th grades (age 8 – 11), the goal of the present research was to provide the first direct investigation of children's metacognition about their self-derivation and how that metacognition might relate to successful self-derivation. The second goal of the present research was to investigate how these relations might differ across development, in the elementary school years of grades 3-5. We hypothesized that if metacognition contributes to successful self-derivation, then we will observe a significant relation between accuracy of metacognitive judgments and self-derivation performance. We hypothesized that if the potential relation between self-derivation and metacognition is strengthened over time, then it may only be apparent in older children.

Method

Participants

Participants were 124 students (66 female, 54 male, 4 unreported) in third grade, 120 students (64 female, 48 male, 8 unreported) in fourth grade, and 86 students (39 female, 39 male, 8 unreported) in fifth grade. All children were students in the same public school in rural southeastern United States. Caregivers/guardians were provided with informed consent forms through children's parent folders as this is the main form of communication between caregivers and the school system. The caregivers of the 320 children included in the sample provided written informed consent for their children to participate prior to the testing session. Children in the same classroom whose caregivers did not provide consent were tested with their classmates, but their data were not included in analyses. Only children whose caregivers returned a signed consent form are included in analyses. The sample was thus the population of children for whom guardians had provided consent for use of their data. For this reason, we did not conduct a power analysis. At the end of the session, all children were given a small token of appreciation for their participation (e.g., pencil). Participating teachers and parents who returned completed demographic questionnaires were thanked with a $10 gift card to a local merchant. All procedures were reviewed and approved by the school board of the participating school system and by the university Institutional Review Board.

Forty percent of children's caregivers returned demographic questionnaires on which they reported on the gender and race/ethnicity of their children. Based on parental report, the sample was 45% Black or African American, 40% non-Hispanic Caucasian, and 11% Hispanic or Latinx. In the year in which the data were collected, 87% of the children in the participating school system were eligible for free or reduced lunch. Of the families who reported caregiver education, 25% had a high school education or less, 31% had some training beyond high school,

16% had a technical or associates degree, 19% had a college bachelor's degree, and 9% has some

training beyond bachelor's degree.

Stimuli

For each of the three grades, stimuli were 16 pairs of true but novel related "stem" facts;

the facts were different for each grade. Each pair of facts could form an integration fact when put

together and subsequently self-derived. Each pair was about a unique topic, developed based on

established curricula and thus appropriate for each grade level. For example, "Venus is the

hottest planet" (Stem 1) and "Another name for Venus is morning star" (Stem 2) form an

integrable pair. From the integration of these facts, participants could derive the knowledge that

"Morning star is another name for the hottest planet." Stimuli changed across grade levels to

align with the grade level appropriate curriculum.

Procedure

Testing took place in children's classrooms over a 45-minute session. The procedure was

administered by two research teams, each consisting of an experimenter and an assistant. Each

research team tested approximately half of all 26 classes.

The session involved three phases (Figure 1). In the first phase children encoded 16 separate

stem facts that were individually projected on a screen at the front of the classroom. Children

then completed a 10-minute unrelated buffer activity. In the second phase, children encoded 16

more stem facts and completed a second unrelated buffer activity for 10-minutes.

At the beginning of the third phase, children were instructed that they were going to

answer some questions about things they may have learned during the experimenters' visit to the

classroom. Children were given a handout of the printed self-derivation questions with spaces for

answers and a scale (1-4) for the feeling-of-knowing judgments for each question (Appendix).

For Feeling-of-Knowing judgments the Likert-type scale ranged from - *definitely will not be*

able to choose the correct answer in multiple choice to 4 - *definitely will be able to choose the correct answer in multiple choice* (Lockl & Schneider, 2002).

The experimenter then oriented participants to the open-ended questions and the metacognitive judgments. First, to demonstrate the low end of the scale, the experimenter provided children with a very difficult question ("When was molybdenum discovered?"). The experimenter explained that because even she, a knowledgeable adult, had no idea what molybdenum is that participants would likely select "1 – not sure at all that I would be able to choose the correct answer in multiple choice." The experimenter then provided an example of an easy question ("What grade are you in?") to prompt children to be very certain they could choose the correct answer in multiple choice if they saw it later and to demonstrate the high end of the scale. She then explained that she would choose "4 – I would definitely be able to choose the correct answer if the same question was seen again in multiple choice format." Following this orientation, children were instructed to read each question, do their best to provide an answer for each question, and then circle the number corresponding to their feeling-of-knowing. Children self-read the questions and recorded their responses in pencil. The questions were presented in one of four different random orders; children seated next to one another had different versions of the test.

After open-ended testing, the same integration questions were presented in forced-choice format through PowerPoint®. The children were then oriented to the Likert-type scale for the confidence judgments. For confidence judgments the Likert-type scale ranged from 1 (not sure at all that I chose the correct answer) to 4 (completely sure I chose the correct answer).The questions were read aloud by the experimenter and children were asked to select the "best answer" from among three options, one of which was correct (33% expected by chance).

Children used individual response devices (a.k.a. "clickers") to provide their responses, which were recorded via TurningPoint® software. The questions were presented in one of four different random orders; each order was used approximately equally often across classrooms. Due to time constraints, we were only able to test recognition of a subset of stem facts (n = 16). Confidence judgments were prompted after each multiple-choice question and children could provide their judgments using the clickers.

Scoring

Children were given 1 point for each self-derivation question answered correctly in open-ended testing. We also calculated scores for children's forced-choice self-derivation performance and forced-choice stem fact memory with each correct answer yielding 1 point. For each of these three question types, participants answered 16 questions.

Data analysis

For each grade, a subset of self-derivation trials was excluded from all analyses because scoring of participants' open-ended responses revealed that a majority of children provided responses to a given question that could not be scored as clear evidence for (or the lack of) self-derivation. This frequently was observed as participants providing what would be the correct answer to a stem fact question in place of the correct answer to a self-derivation question. These responses suggest that participants were answering in the right domain but it was unclear whether they had successfully self-derived through integration. For all analyses, we removed 2 trials from all 3rd graders, 1 trial from all 4th graders, and 4 trials from all 5th graders. Because not all 16 questions were included in the analyses for each grade, we converted all self-derivation and stem fact memory scores to proportions.

To measure participants' relative metacognitive accuracy, we calculated three gamma correlations for each participant using the Hmisc package in R (Harrell, 2020). Gamma

correlations reflect the relative accuracy of metacognitive judgments by comparing pairs of item scores and their judgments based on whether they agree or disagree (Nelson, 1984). This correlation is a proportion of concordant judgments over total judgments made. For example, a high confidence rating for an item answered correctly would be an agreement whereas a high confidence rating for an item answered incorrectly would be a disagreement. Participants with more agreements than disagreements are more metacognitively accurate than those who have more disagreements or do not vary in their judgments at all. Gamma can range from -1 to 1. We calculated a gamma correlation between the feeling of knowing judgments elicited at open-ended test for self-derivation and subsequent forced choice self-derivation performance, forced choice self-derivation performance and their respective confidence judgments, and forced choice stem fact memory and respective confidence judgments.

Each gamma correlation was then correlated to open-ended self-derivation performance, forced-choice self-derivation performance, and forced-choice stem fact memory. The correlation between gamma and performance reflects the strength of the relation between self-derivation or stem fact memory and the accuracy of associated metacognitive judgments.

Because self-derivation performance and metacognitive judgments across multiple trials were nested within each child and children were nested within grades, we used multi-level modeling (MLM) to test relations between self-derivation and metacognitive judgment. Multi-level modeling controls for the nesting of data within individuals and allows for individual intercepts. Specifically, it allows for predictors at both the trial and person levels. It also takes into account the interdependency of multiple observations per person (e.g., 12-15 trials for each child), correcting for the biases in parameter estimates resulting from dependency of the observations (Raudenbush & Bryk, 2002; Wright, 1998). Multi-level modeling is also amenable

to missing data or unequal numbers of trials across participants, both of which were present in the current data. Specific information on missing data for each grade level are provided in the Appendix. As reflected there, there was more missing data for the earlier than later grades.

<div align="center">**Results**</div>

The results are presented in four sections: self-derivation and stem-fact performance, metacognitive performance, the Pearson correlations between self-derivation and metacognition, and multilevel models.

Self-derivation and stem-fact performance

Average proportion correct for open-ended self-derivation, forced choice self-derivation, and forced choice stem-fact memory performance by grade is reported in Table 1. A one-way ANOVA revealed a difference in average open-ended self-derivation across grades $F(2, 312) = 10.673$, $p < .001$, $\eta_p^2 = .064$, and Bonferroni corrected pairwise comparisons reflected that both 4th and 5th graders performed significantly higher than 3rd graders; 4th and 5th graders did not differ from one another. A one way ANOVA revealed the same pattern of results for forced choice self-derivation performance $F(2, 310) = 23.277$, $p < .001$, $\eta_p^2 = .131$. Again, 4th and 5th grade students performed significantly higher than 3rd grade students after Bonferroni correction, and did not differ from one another. The main effect of grade supports age-related change in self-derivation performance as older children were correct on more self-derivation questions than younger children. Overall, open-ended self-derivation performance was low and thus has little variability. For this reason, all further analyses utilize only forced choice data.

A one-way ANOVA revealed a difference in forced-choice stem-fact performance across grades $F(2, 278) = 6.956$, $p = .001$, $\eta_p^2 = .048$. Specifically, 4th graders' stem-fact performance was significantly higher than that of both 3rd and 5th graders after Bonferroni corrections for multiple comparisons. Third and 5th grade students did not differ from each other in their forced-

choice stem-fact performance. Significantly higher stem fact memory also lends support to higher self-derivation performance in 4th graders because stem facts are necessary, but not sufficient for successful self-derivation. Relatedly, 5th graders' lower stem fact memory than that of 4th graders provides evidence for why they did not demonstrate higher self-derivation performance than 4th graders.

Metacognitive performance

Grade-level mean metacognitive judgments are reported in Table 1. There were no significant differences in average judgment across grade for any of the three judgment types: (a) feelings-of-knowing predicting subsequent forced-choice self-derivation performance: $F(2, 310) = 0.454$, $p = .635$, $\eta_p^2 = .003$; (b) confidence judgments made after forced-choice self-derivation performance: $F(2, 293) = 0.899$, $p = .408$, $\eta_p^2 = .006$; and (c) confidence judgments made after forced-choice stem-fact performance: $F(2, 245) = 0.889$, $p = .412$, $\eta_p^2 = .007$. These findings reflect that there were no age-related increases in metacognitive judgments. Metacognition about self-derivation was not different across grade levels.

Mean gamma correlations are reported for each performance-judgment combination by grade in Table 1. The metacognitive literature reflects that a gamma correlation of .5 would be considered moderate as it indicates that participants' judgments agreed with their actual performance on approximately half of all trials. A negative gamma reflects that an individual's judgments tend to be in opposition to actual performance. For example, a participant with a tendency to have very low confidence on correct trials but not on incorrect trials would likely have a negative gamma correlation. Overall, participants exhibited low relative accuracy of their metacognitive judgments with averages ranging from $G = .16$ to $G = .40$. Further, we did not observe a significant difference in metacognitive accuracy across grades for any of the three

gamma correlations: (a) gamma correlation between feelings of knowing and forced-choice self-derivation performance: $F(2, 278) = 1.476$, $p = .230$, $\eta_p^2 = .011$; (b) gamma correlation between forced-choice self-derivation performance and subsequent confidence judgments: $F(2, 281) = 2.611$, $p = .075$, $\eta_p^2 = .018$; and (c) gamma correlation between forced-choice stem-fact performance and subsequent confidence judgments: $F(2, 206) = 1.488$, $p = .228$, $\eta_p^2 = .014$. This reflects that accuracy of metacognition also did not differ across grade levels.

We also analyzed whether each gamma correlation was significantly different from zero. Because we did not observe significant differences in gamma correlations across grades, we collapsed across grade. All three gamma correlations were significantly different from zero: (a) gamma correlation between feelings of knowing and forced-choice self-derivation performance: $t(280) = 6.411$, $p < .001$, $d = .38$; (b) gamma correlation between forced-choice self-derivation performance and subsequent confidence judgments: $t(284) = 8.701$, $p < .001$, $d = .52$; and (c) gamma correlation between forced-choice stem-fact performance and subsequent confidence judgments $t(209) = 6.600$, $p < .001$, $d = .46$. Gamma correlations significantly different from zero demonstrate that there is a significant relation between metacognitive judgments and task performance about which the judgments were made.

Correlations between self-derivation and metacognition

Correlations between self-derivation performance and stem fact memory are presented in Table 2. Relations between average metacognitive judgments are reported in Table 3. Correlations between self-derivation measures and metacognitive measures, including gamma, are presented in Table 3. Because we observed no clear pattern of grade-related differences in either self-derivation measures or metacognitive measures, all correlations are presented collapsed across grade levels.

Children's self-derivation performance was significantly related to their stem fact memory, reflecting that knowledge of stem facts is necessary for successful self-derivation, $r(281) = .44$, $p < .001$.

Feeling of knowing judgments were not related to confidence judgments about self-derivation performance, $r(290) = .11$, $p = .06$. They also were not related to confidence judgments about stem fact performance, $r(245) = .08$, $p = .23$. However, confidence judgments about self-derivation and stem fact performance were related to each other, $r(248) = .71$, $p < .001$. Participants who were highly confident in their self-derivation performance also tended to be confident in their stem fact memory.

Accuracy of self-derivation confidence judgments was related to both forced-choice self-derivation performance, $r(284) = .18$, $p = .003$. This reflects that those with more accurate metacognition are also more successful at self-deriving, suggesting metacognition may support successful generation of new knowledge. Self-derivation confidence judgment accuracy was also related to stem fact memory, $r(254) = .23$, $p < .001$. Finally, accuracy of self-derivation confidence judgments was significantly correlated with accuracy of feeling of knowing judgments, $r(258) = .17$, $p = .007$. This relation suggests that metacognitive accuracy is relatively consistent within individuals. There were no other significant correlations with metacognitive accuracy and self-derivation or stem fact performance.

Models

To predict the variance in self-derivation explained by metacognition and account for the nested structure of trials within participants across 3 grades, we used multilevel modeling. We first tested 2 null models to determine whether there was variance at both levels of the model. First, we constructed a null model with individual forced-choice self-derivation trials at Level 1

and grade at Level 2 to determine the variance between grades. The intraclass correlation for this null model reflected only 0.04% of the variance between grades. In a second null model, we examined whether there was sufficient variance between participants; again, the intraclass correlation reflected limited variance of 0.70% at this level. These analyses suggest that though the current study has a nested structure, the majority of the observed variability is at the level of the individual forced-choice self-derivation trial rather than at the level of the participant or grade. These findings reflect that there were no age-related differences in self-derivation or in how metacognition is related to self-derivation. Variability primarily at the level of the individual trial suggests that variables like whether participants recalled the stem facts necessary to self-derive on a given trial explained the differences observed in performance.

Discussion

The present research had two primary purposes. The first was to investigate the potential relation between children's self-derivation performance and metacognition. The second primary purpose was to examine developmental change in the relation between self-derivation and metacognition.

We first examined self-derivation performance across grades 3 through 5. There was a significant difference in the open-ended self-derivation performance across grade level such that 4th and 5th graders performed significantly higher than 3rd graders. However, the 4th and 5th grade students were only correct on 15% and 14% of trials, respectively. This low performance in open-ended testing has been observed in previous classroom-based self-derivation research (Esposito & Bauer, 2017; Varga et al., 2020). Because such low performance leaves little variability to examine, it has not been the focus of analysis and was not analyzed in the present research. As we expected, there was an age-related increase in forced-choice self-derivation performance. Both 4th and 5th grade students performed significantly better than 3rd graders.

There was no difference between the 4th and 5th graders, perhaps suggesting a slowing of the rate of change in self-derivation performance. It is also likely that variability existing largely at the level of the individual trial is why there is significant difference between older students. Alternatively, we may not have observed a difference between the upper grades because the stimuli were created based on the grade-level curriculum and may have been optimally challenging for the older students.

We also observed a significant difference in forced-choice stem fact memory. As we might expect based on higher self-derivation performance in the 4th graders, they remembered more stem facts than their peers. Specifically, they remembered 69% of stem facts whereas the 3rd and 5th grade students remembered only 57% each. Importantly, 5th graders' significantly lower stem fact memory may be an issue of power. Fifth graders were the smallest group in our sample, rendering them most sensitive to data loss, especially to loss of stem fact performance data. The forced-choice stem fact questions were the last questions children answered. Thus, the likelihood that participants abandoned their participation in our fact "game" was at its peak.

To investigate children's metacognition about their self-derivation, we tasked children with providing metacognitive judgments after answering open-ended self-derivation questions as well as after forced-choice self-derivation and stem fact questions. We examined change in both the average judgment and the accuracy of the metacognitive judgments. The majority of participants in the sample used two or more points on 4-point (1 through 4) the metacognitive judgment scales. However, across the entire sample and all judgment types, participants provided an average judgment of 3. Feeling-of-knowing judgments following open-ended self-derivation questions serve as a child's predictions of how likely she is to choose the correct answer when presented with the same questions in multiple choice format. There was no difference in

children's FOKs across grade level. On average, participants selected "2 – I probably will not be able to choose the correct answer" suggesting low certainty in future performance. The same pattern was true for the confidence judgments following forced-choice self-derivation and stem fact questions. However, participants tended to rate their confidence as "3 – Pretty sure" that they chose the correct answer. This increase in the average judgment, relative to FOKs, is fitting when considering that the forced-choice self-derivation questions had been seen before in open-ended and that stem facts were directly taught to the participants. This also highlights a unique insight into self-derivation: children in all three grades were equally confident in their self-derivation and stem fact performance even though only the latter had been presented to them explicitly.

Because metacognition can positively influence cognition when metacognitive monitoring is accurate, it is possible that self-derivation is related to accurate metacognitive judgments. To inspect the accuracy of participants' metacognitive judgments, we calculated gamma correlations, a commonly used index, for feeling-of-knowing judgments, post self-derivation question confidence, and post stem fact question confidence. On average, for 3rd graders $G = .16$ reflecting that our youngest participants accurately predicted their forced-choice self-derivation performance on only 16% of trials. Accuracy increased in both 4th ($G = .22$) and 5th graders ($G = .32$) but not significantly. This pattern of non-significant increases in metacognitive accuracy in the 4th and 5th graders was observed for confidence judgments in both forced-choice self-derivation and stem fact memory. Crucially, metacognitive accuracy as it relates to self-derivation is considerably lower than what is typically observed of children in the 3rd to 5th grade age range, approximately .50 to .70 depending on the task that participants complete. This may be due simply to the number of trials and judgments children were asked to

complete in a relatively short time period. Low metacognitive accuracy may also be related to the challenging nature of self-deriving through integration of separate episodes of learning. In studies of metacognition, children typically make judgments about less challenging stimuli such as word pairs.

Further explication of the relation between self-derivation and metacognition requires examination of the relation between gamma, our index of metacognitive accuracy, and self-derivation performance. We first collapsed across all grades and correlated all participants' forced-choice outcomes with the three calculated gammas. Across all participants, there was a significant relation between forced-choice self-derivation and FCI-Gamma, reflecting that one's level of performance was related to the accuracy of their confidence judgments. Those who were more successful at generating new knowledge also were more metacognitively accurate. Furthermore, FCI-Gamma was related to FOK-Gamma suggesting that participants' metacognition was consistently accurate across different judgment type. To examine change in these relations over time, we separated participants by grade and conducted correlations. When separated, only 3rd graders' self-derivation performance was related to their FCI-Gamma. For both 3rd and 4th grades, FCI-Gamma was significantly related to forced-choice stem performance. This suggests that participants who are more metacognitively accurate remember more stem facts, possibly due to the information about one's own learning gained by monitoring.

Investigating developmental change in the hypothesized relation between metacognitive accuracy and self-derivation was a driving motivation for the present research. In our sample, we did not observe robust age-related change in either self-derivation or metacognition. We did find evidence suggesting that self-derivation and metacognition are related. Contrary to our hypothesis, we did not observe that this relation changes across ages 8 years to 11 years. Despite

the marked changes typically observed across development in laboratory studies of metacognition, it is plausible that the pattern observed in the present work mirrors what one would find "in the wild" because participants in our study were tested in their home classrooms in a format not unlike the methods frequently employed by their teachers. We must also consider the information children learned in each grade level. The stem fact pairs were developed to be appropriate for each grade level's curriculum. Had all participants across the sample learned the same information, it is likely that 5th grade students would have exhibited higher confidence than 3rd or 4th graders due to the relatively easier material. Furthermore, Lockl and Schneider (2002) found little change in the feeling of knowing accuracy participants age 7 to 10 years. These participants were tasked with recognizing words and their associated images after providing feeling-of-knowing judgments. Overall, the sample's FOK accuracy was significantly above zero but functionally rather low (Lockl & Schneider, 2002). Given the relatively simple nature of recognizing word-picture pairs, it is logical to conclude that metacognitive accuracy is rather sensitive to task difficulty. Evidence of this in the present research is supported by the multilevel models reflecting that most of the variance was at the level of the individual trial; participants demonstrated varied confidence based on the perceived difficulty of individual trials.

Though our hypotheses about the relation between self-derivation and metacognition were supported only in part, the present study made important contributions. Little is known about metacognition for knowledge gained through productive processes. The extant evidence suggesting metacognition facilitates the success of processes has come from only studies of adults (Smith & Squire, 2005; Varga & Bauer, 2017a). This study provides additional evidence to our understanding, as well as a developmental perspective. Second, the purpose of the current research was not to directly compare metacognition for stem facts to that of self-derived

knowledge. However, the findings provide important insight into the epistemological status of that which is directly learned (i.e., stem facts) compared to that which is self-derived. Metacognitive judgments and accuracy did not differ for self-derivation versus stem fact performance. Participants appear to learn and monitor self-derived knowledge just as well as directly learned stem facts. A study of 8-year-olds' source memory for self-derived knowledge also demonstrates that children treat self-derived knowledge and that which is directly learned in largely the same manner (Dugan & Bauer, 2020). Children were tasked with nominating where they learned explicitly-taught facts, self-derived facts, and stem facts. Participants were accurate when selecting which story a stem fact came from. When selecting the sources for explicitly-taught and self-derived information, children tended to select the story that contained the opportunity to gain knowledge.

In summary, we examined developmental change in self-derivation and metacognition. We did not find compelling evidence for age-related change, particularly for the metacognitive measures. We tested for a relation between self-derivation and metacognition and found that only when all participants are analyzed together, accuracy of confidence judgments is related to forced-choice self-derivation performance. Thus, we have the first evidence to suggest that metacognition is related to, and potentially facilitates, self-derivation among 3rd through 5th grade children. However, to get a clearer picture of any change in the relation between self-derivation and metacognition, a longitudinal investigation is critical.

Limitations and Directions for Future Research

The present research is not without limitations. As can be expected of data collection with large groups of children in their familiar classroom environment, missing data were frequent across our sample. This was particularly true of the younger children (data reflected in

the Appendix). In order to retain statistical power, all assenting children were included in our analyses. Had we set a criteria for number of trials completed, we may have artificially biased our sample. This is especially when considering that metacognitive accuracy is calculated on a trial by trial basis. We did not want to introduce any systematic differences in metacognitive accuracy by including only children who completed a high proportion of trials.

The cross-sectional design of the present research afforded an investigation of metacognition and self-derivation in a cross-sectional sample of 3rd through 5th grade students. The design also placed important constraints on our understanding of children's metacognition about self-derivation. To avoid influencing children's learning of the separate stem facts, we could only elicit metacognitive judgments during the test phases. Thus, we were limited to conjecture in understanding children's monitoring of their learning as it was occurring. However, encoding facts without explicitly providing metacognitive judgments also reflects any intrinsic or endogenous metacognition children were "naturally" engaging in during the learning phases of the study.

Results of the present research reflect that metacognition and self-derivation are related in elementary school students, at least as tested in this study. Based on prior research with feelings-of-knowing, we utilized a 4-point Likert-type scale (Lockl & Schneider, 2002). A Likert-type scale of more than 4 points may have provided the opportunity to observe greater variability in children's judgments and their relative metacognitive accuracy. If there were more points on the metacognitive scale and children continued to utilize nearly the entire scale, then there lies the possibility that greater variability in metacognitive judgments would contribute to different, more robust patterns of metacognitive accuracy.

Future research should employ Judgments of Learning to better understand how children engage in the early steps of ERISS (Bauer & Varga, 2017), encoding, reactivation, and integration. A pattern of higher JOLs to the second stem fact than the first could suggest a level of understanding that the stem facts are integrable. Higher JOLs for either fact in a pair may also predict subsequent self-derivation. Furthermore, investigation of metacognitive control in the context of self-derivation is necessary to gain a holistic view of the ways in which metacognition impacts self-derivation. Further research on metacognition, particularly monitoring during encoding of related facts, is necessary to draw broader conclusions about the role of metacognition in successful self-derivation.

Conclusion

In conclusion, the present research provides insight into the relation between self-derivation and metacognition and into metacognition for knowledge that is derived internally rather than taught directly. It is clear that more research investigating metacognition as it relates to the process of self-derivation is necessary. A longitudinal perspective would be especially beneficial as it could help elucidate age-related change that is typically observed in studies of self-derivation and of children's metacognition. Importantly, the present data reflect that for children in grades 3 through 5 the accuracy of their confidence judgments about self-derivation is related to self-derivation performance. This is the first evidence that metacognition may support self-derivation in children. The present research represents a meaningful foray into understanding how children think about their self-derived knowledge.

References

Bauer, P. J., Blue, S. N., Xu, A., & Esposito, A. G. (2017). Productive extension of semantic memory in school-age children: Relations with reading comprehension and deployment of cognitive resources. *Developmental Psychology*, *52*(7), 1024.

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., & Larkina, M. L. (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., & Varga, N. L. (2017). Similarity and Deviation in Event Segmentation and Memory Integration: Commentary on Richmond, Gold, & Zacks. *Journal of Applied Research in Memory and Cognition*, *6*(2), 124-128.

Besner, N. R., & Son, L. K. (2007). Underlying mechanisms of initial feelings of knowing in children. *Scandinavian Journal of Psychology*, *48*(6), 449-457.

Dugan, J. A., & Bauer, P. J. (2020) The process of self-derivation: Insights from 8-year-olds' eye movements, memory, and source attributions. *Manuscript in preparation.*

Dunlosky, J., & Nelson, T. O. (1994). Does the sensitivity of judgments of learning (JOLs) to the effects of various study activities depend on when the JOLs occur?. *Journal of Memory and Language*, *33*(4), 545-565.

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.

Finn, B. & Metcalfe, J. (2014). Overconfidence in Children's Multi-Trial Judgments of Learning. *Learning and Instruction, 32,* 1-9

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*(10), 906.

Goswami, U. (2011). Inductive and deductive reasoning. In U. Goswami (Ed.), Childhood cognitive development (pp. 399–419). Oxford, UK: Wiley–Blackwell. doi.org/10.1016/j.tics.2010.10.

Hacker, D. J., Dunlosky, J., & Graesser, A. C. (Eds.). (2009). Handbook of metacognition in education. Routledge.

Harrell, F. E. (2020). Hmisc: Harrell Miscellaneous. R package version 4.4-0. https://CRAN.R-project.org/package=Hmisc

Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology*, *56*(4), 208.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*, 349-370.

Koriat, A. (2012). The relationships between monitoring, regulation, and performance: Commentary. *Learning and Instruction*, *22*, 296-298.

Lockl, K., & Schneider, W. (2002). Developmental trends in children's feeling-of-knowing judgements. *International Journal of Behavioral Development*, *26*(4), 327-333.

Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review*, *15*(1), 174-179.

Metcalfe, J. & Finn, B. (2013). Metacognition and control of study choice in children. *Metacognition and Learning, 8*(1), 19-46.

Nelson, T.O. & Narens, L. (1990). Metamemory: A theoretical framework and some new

findings. In G.H. Bower (Ed). The Psychology of Learning and Motivation, 26, 125-173.

New York: Academic Press

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing

predictions. *Psychological Bulletin, 95*(1), 109–133. https://doi.org/10.1037/0033-

2909.95.1.109

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data

analysis methods* (Vol. 1). sage.

Roebers, C. M. & Spiess, M. (2017). The development of metacognitive monitoring and control

in second graders: A short-term longitudinal study. *Journal of Cognition and

Development, 18*, 110-128.

Roebers, C. M., Mayer, B., Steiner, M., Bayard, N., & Van Loon, M. (2019). The role of

children`s metacognitive experiences for cue utilization and monitoring accuracy: A

longitudinal study. *Developmental Psychology*. Online First Publication. Doi:

10.1037/dev0000776

Roebers, C. M., von der Linden, N., Schneider, W., & Howie, P. (2007). Children's

metamemorial judgments in an event recall task. *Journal of Experimental Child

Psychology, 97*, 117 - 137.

Schneider, W. (1985). Developmental trends in the metamemory-memory behavior relationship:

An integrative review. In D. Forrest-Pressley, G. E. MacKinnon, & T. G. Waller (Eds.),

Metacognition, Cognition, and Human Performance Vol 1 (p. 57-109). Academic Press;

Boston.

Schneider, W., & Löffler, E. (2016). The development of metacognitive knowledge in children and adolescents. In J. Dunlosky & S. K. Tauber (Eds.), *Oxford library of psychology. The Oxford handbook of metamemory* (p. 491–518). Oxford University Press

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology*, *19*(4), 460-475.

Smith, C., & Squire, L. R. (2005). Declarative memory, awareness, and transitive inference. *The Journal of Neuroscience, 25(*44), 10138-10146.

Tsalas, N., Paulus, M., & Sodian, B. (2015). Developmental changes and the effect of self-generated feedback in metacognitive controlled spacing strategies in 7-year-olds, 10-year-olds, and adults. *Journal of Experimental Child Psychology*, *132*, 140-154.

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. (2020). Cognitive correlates of memory integration across development: Explaining variability in an educationally relevant phenomenon. *Journal of Experimental Psychology: General*, *148*(4), 739.

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. (DOI: 10.1016/j.jecp.2013.01.008)

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.

Visé, M., & Schneider, W., (2000) Determinants of performance prediction in kindergarten and school children: the importance of metacognitive and motivational factors. *Zeitschrift fur Entwicklungspsychologie und Padagogische Psychologie, 32*(2), 51-58.

Wright, D. B. (1998). Modelling clustered data in autobiographical memory research: The multilevel approach. *Applied Cognitive Psychology, 12*, 339–357.

Table 1

Self-derivation and Metacognitive Measure Descriptives in Mean (SD)

Measure	Grade 3	Grade 4	Grade 5
OE Self-derivation Proportion Correct	.09 (.12)	.16 (.17)	.18 (.17)
FC Self-derivation Proportion Correct	.46 (.18)	.62 (.20)	.59 (.18)
FC Stem Fact Proportion Correct	.57 (.22)	.69 (.22)	.57 (.33
Average Feeling-of-Knowing	2.19 (.66)	2.15 (.65)	2.24 (.60)
Average Confidence in FC Self-derivation	2.94 (.64)	3.02 (.54)	2.91 (.55)
Average Confidence in FC Stem Fact Memory	3.25 (.67)	3.14 (.64)	3.10 (.71)
Average FOK-FCSD Gamma	.16 (.51)	.22 (.63)	.31 (.61)
Average FCSD-CJ Gamma	.20 (.56)	.38 (.55)	.30 (.63)
Average FCS-CJ Gamma	.25 (.67)	.40 (.63)	.22 (.74)

Table 2

Correlations between Self-derivation and Stem Fact Performance Measures

	OE Self-derivation Proportion Correct	FC Self-derivation Proportion Correct	FC Stem Fact Memory Proportion Correct
OE Self-derivation Proportion Correct	1	.681[**]	.371[**]
FC Self-derivation Proportion Correct		1	.440[**]
FC Stem Fact Memory Proportion Correct			1

Note. **. Correlation is significant at the 0.01 level (2-tailed

Table 3

Correlations between Metacognitive Judgments

	Mean FOK	Mean FCSD-CJ	Mean FCS-CJ
Mean FOK	1	0.112	0.077
Mean FCSD-CJ		1	.713[**]
Mean FCS-CJ			1

Note. **. Correlation is significant at the 0.01 level (2-tailed).

Table 4

Relations between Self-derivation and Metacognitive Measures

	FC Self-derivation Proportion Correct	FC Stem Fact Memory Proportion Correct	FOK Accuracy	FCSD-CJ Accuracy	FCS-CJ Accuracy
FC Self-derivation Proportion Correct	1	.440**	0.086	.178**	0.116
FC Stem Fact Memory Proportion Correct		1	0.022	.229**	0.077
FOK Accuracy			1	.168**	-0.003
FCSD-CJ Accuracy				1	0.107
FCS-CJ Accuracy					1

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 1. Schematic of Procedure in Classroom

Encoding Phase 1:
16 Stem Facts

Ex. Venus is the hottest planet.

Buffer Activity

Encoding Phase 2:
16 Stem Facts

Ex. Another name for Venus is Morningstar.

Buffer Activity

Test Phase 1: Open-ended
16 Open-ended Self-derivation questions

Ex. What is another name for the hottest planet?   1 2 3 4

Test Phase 2: Forced-choice
16 Forced-choice Self-derivation questions

Ex. What is another name for the hottest planet?

a) Morningstar
b) Hot star
c) Heat wave

How sure are you that you chose the correct answer?

1 – not sure at all
2 – a little sure
3 – mostly sure
4 – definitely sure

16 Forced-choice stem fact questions

Ex. What is the hottest planet?

a) Earth
b) Jupiter
c) Venus

How sure are you that you chose the correct answer?

1 2 3 4

Appendix

Table 1

Datapoints missing from Multilevel Models

Grade	Total Missing Datapoints	FOK	FC Self-derivation	FCSD-CJ
3	746	28	383	335
4	421	0	291	130
5	198	0	125	73

Multilevel Model Equations

Null Model

Level 1: Self-derivation$_{it}$ = β_{0it} + r_{it}

Level 2: β_{0i} = γ_{00} + u_{0i}

B$_{0i}$ is the mean self-derivation for participant i, γ_{00} is the overall mean self-derivation of the sample, and u_{0i} represented the degree to which individuals vary from the sample as a whole.

Model 1

The random coefficients regression models equations used to test whether metacognition predicted self-derivation were:

Level 1: Self-derivation$_{it}$ = β_{0it} + β_{1it}(metacognition) + r_{it}

Level 2: β_{0i} = γ_{00} + u_{0i}

β_{1i} = γ_{10} + u_{1i}

The slope coefficient β_{1it} represented the associated change in self-derivation associated with change in metacognitive 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 self-derivation for the sample. Further, γ_{10} corresponded to the effect of metacognition on self-derivation and u_{0i} and u_{1i} represent the degree to which individuals vary from the sample as a whole.

Study 3

Metacognition during self-derivation: A longitudinal analysis

Jessica A. Dugan & Patricia J. Bauer

Emory University

Metacognition during self-derivation: a longitudinal analysis

"You learn something new every day." There is perhaps no time when this is truer than in the elementary school years as young learners regularly encounter new information. Not only are learning opportunities abundant, there are many ways to gain new knowledge. Several of these ways are easily recognized as direct means for acquiring knowledge: reading a book, listening to a classroom lecture, or interacting with a museum exhibit. Importantly, learning also occurs through productive processes that allow one to extend knowledge to learn something new. Productive processes include analogy, induction, deduction, and the focus of the current research – self-derivation through memory integration. Learning opportunities are also supported by other cognitive abilities. Metacognition, or thinking about thinking, is the suite of processes that involves the monitoring and control of one's cognition (Flavell, 1979; Nelson & Narens, 1990). Metacognition is of particular importance because it develops in childhood and may directly impact productive processes as they unfold. Metacognition and direct means of learning have been the focus of much research over the last century, yet little is known of children's metacognition during self-derivation. To bridge the gap in the current understanding of the relation between children's metacognition and self-derivation and whether there is intraindividual stability in this relation, the present research was a longitudinal investigation of children's metacognition about their self-derivation performance over 3 years.

Perhaps one of the most critical tasks to span development is the building of a knowledge base. Knowledge base development is aided greatly by productive processes, such as analogy, deduction, transitive inference (see Goswami, 2011 for a review), and self-derivation, that allow one to go beyond what is learned directly to gain new knowledge. Self-derivation is a pervasive process that unfolds over multiple subprocesses and occurs through the integration of separate but related episodes of new learning (Bauer & Varga, 2017). For example, in a study where

children learned pairs of related facts (i.e. "stem" facts) from story passages, children learned from one story the true but novel fact that "dolphins live in groups called pods" (Stem 1). From a story later in the same study session, children learned the related fact "dolphins talk by clicking and squeaking" (Stem 2). When faced with the question "How does a pod talk?" children who integrate facts across the stories can then derive the new knowledge that "pods talk by clicking and squeaking" (Bauer & Larkina, 2017; Bauer & San Souci, 2010).

Of all productive processes, self-derivation is the focus of the present research for several reasons. In the self-derivation paradigm, participants learn new factual information from ecologically relevant domains such as history, geography, literature, and physics, to name but a few. Knowledge self-derived through integration is rapidly incorporated into the knowledge base (Bauer & Jackson, 2015) and is retained for at least one week after initial testing (Dugan & Bauer, 2020; Varga & Bauer, 2013; Varga, Stewart, & Bauer, 2016). In addition to the materials of self-derivation being information one might learn in a classroom, self-derivation performance is also significantly related to academic achievement in children and adults (Varga, Esposito, & Bauer, 2019). Several studies also demonstrate that there is significant age-related change in self-derivation performance (Bauer, Cronin-Golomb, Porter, Jaganjac, & Miller, 2020; Bauer & Larkina, 2017; Varga & Bauer, 2017a). With respect to relations with metacognition, what makes self-derivation a particularly compelling process is that self-derived knowledge arises in the mind of the learner instead of coming directly from experiences in the world. Furthermore, the multistep nature of self-derivation means there are several points along the way that metacognition may impact. These qualities mark self-derivation as a unique opportunity to investigate how one's metacognition may support the generation of new knowledge and whether such a relation may change over development.

Previous work makes clear that there is age-related change in the success one experiences at self-deriving new knowledge. Self-derivation has been observed in children as young as 4-years old (Bauer & San Souci, 2010), across childhood (Bauer & Larkina, 2017; Dugan & Bauer, 2020; Esposito & Bauer, 2017), and in young adult college students (Varga & Bauer, 2017a; 2017b). To illustrate, Bauer and Larkina (2017) conducted a laboratory study with children ages 4-, 6-, and 8-years-old wherein children learned pairs of related stem facts presented in story passages. Self-derivation performance increased across the three observed age groups such that 4-year-olds successfully self-derived new knowledge on 13% of trials and 8-year-olds were successful on 75% of trials (Bauer & Larkina, 2017). Importantly, when children are provided with only one of the stem facts from a pair, they do not go on to self-derive new knowledge (Bauer & Larkina, 2017; Bauer & San Souci, 2010). Not only is there age-related change in self-derivation performance, an investigation of the cognitive correlates (e.g., verbal comprehension) of self-derivation makes clear that the individual variability observed in self-derivation is systematic (Varga, Esposito, & Bauer, 2019).

Self-derivation also is observed in the classroom. Esposito and Bauer (2017) used the same story passage paradigm as previous research (Bauer & Larkina, 2017; Bauer & San Souci, 2010; Varga & Bauer, 2013) in classrooms of students grades K through 3. Though performance in the classroom was lower than is typically observed in the laboratory, a similar pattern of age-related increases in self-derivation performance was observed (Esposito & Bauer, 2017). Additionally, there is individual variability such that even the oldest children ranged from 0% to 100% correct in open-ended testing and 25% to 100% correct in forced-choice testing. Lower performance compared to laboratory research is likely attributable to the common distractors one might face in a classroom: technology, peers talking, or even the lunch menu. These studies

demonstrate clear evidence of age-related change in self-derivation performance, both in the laboratory and in the classroom, as well as pervasive individual differences in performance.

Individual differences in cognitive abilities may contribute to both age-related change and the individual variability in self-derivation success. An especially salient candidate ability is metacognition. Consider that knowledge that is self-derived is a product of one's own mind rather than a direct experience. Metacognition is "thinking about thinking" or the monitoring and control one exerts over cognition (Flavell, 1979; Nelson & Narens, 1990). Because metacognition involves attending to and controlling one's cognition, it may contribute to variability in successful generation of new knowledge across development and between individuals. Metacognitive monitoring provides knowledge about cognition. In the context of self-derivation, one learns novel facts. When later learning related facts, including the second stem in a pair, accurate monitoring may increase the likelihood that the shared information may act as a cue to reactivate Stem 1. It is possible that the learner may then try to integrate the separate but related facts and go on to self-derive new knowledge. Throughout this process of self-deriving new knowledge, there are many opportunities for metacognition to contribute to the variability that is consistently observed in self-derivation performance.

Metacognitive monitoring is not just a process for individuals to gather information about their own learning: it is also a tool often utilized by researchers to understand the role of metacognition in another task. Monitoring involves, for example, attending to cues about the difficulty of a test question or whether one will be able to remember something later. Metacognitive judgments are indices of how well one thinks she will remember a fact later (Judgment of Learning), predictions of whether the correct answer would be recognized from amongst distractors in a multiple choice test (Feeling of Knowing), or even of how certain she is

of her performance after taking a test (Confidence Judgments). Studies using metacognitive judgments often have participants provide their judgments on a scale of 0 to 100 with 0 being lowest confidence and 100 being highest confidence. For example, in a study conducted by Finn and Metcalfe (2014), children in grades 3 and 5 learned definitions of words and then made judgments of learning as percentages between 0% and 100%. These judgments are considered accurate if they align well with a participant's actual performance. Gamma correlations are effectively the proportion of metacognitive judgments that are in concordance, rather than discordance, with actual task performance. Ranging between -1 and 1, higher gamma correlations indicate higher metacognitive accuracy. When gamma is high and is also significantly correlated with performance on a task like self-derivation, it suggests that accurate metacognition supports successful task performance.

To date there has been no systematic investigation of metacognition in the context of self-derivation. However, it is logical to hypothesize that metacognition may relate to and even influence self-derivation. First, self-derivation is a process that unfolds over 5 steps (Bauer & Varga, 2017), providing multiple opportunities at which to monitor and control one's cognition. Encoding, the first step in the ERISS model of self-derivation, is likely a critical time for metacognitive monitoring. Attending closely to facts could lead to better memory for the stems necessary for successfully creating new knowledge. It may also aid in Reactivation as monitoring may increase the likelihood of considering what else might be known about or related to new information, such as Stem 2 of an integrable pair of facts. Integration and Selection, the third and fourth steps of ERISS, could be impacted by actively attempting to put together all the information about a particular topic and considering what information is relevant at a given time.

Finally, metacognition at any one of these steps may increase success at Self-derivation and give learners knowledge about their own capacity for generating new knowledge.

Just as there is variability in self-derivation, so too is there persistent variability in metacognition (Roebers & Spiess, 2017; Schneider, 1985; Schneider & Lockl, 2002). Preschoolers and kindergarteners are significantly less accurate than children in second to third grades at predicting future memory performance as they tend to be overconfident in their memories (Vise & Schneider, 2000). A marked shift in the accuracy of metacognition appears around age 7- to 8-years (Metcalfe & Finn, 2013; Roebers & Spiess, 2017; Tsalas, Paulus, & Sodian, 2015). Roebers and Spiess (2017) tracked the development of metacognitive monitoring for spelling in 7-year-olds. Over an 8-month period, children significantly improved in their monitoring of the moderately difficult task. Another longitudinal study of 7- to 11-year-olds further highlights the meaningful change in metacognitive accuracy observed over the course of just a school year (Roebers et al., 2019). By the final study visit, 11-year-olds had a mean gamma correlation of .58. This nearly 10% increase in accuracy from the moderate gamma correlation of .50 elevated their metacognitive performance to adult-like levels (Roebers et al., 2019). These age-related increases in metacognition appear to mirror the age-related increases observed in self-derivation. A longitudinal investigation is necessary to elucidate the contribution metacognition makes toward both the age-related and individual variability observed in self-derivation performance.

Present research

Little is currently known about the relation between self-derivation and metacognition. Both metacognition and self-derivation undergo substantial development in the elementary school years such that older children are both more accurate in their metacognition and self-

derive at higher levels than younger children (Bauer & Larkina, 2017; Esposito & Bauer, 2017; Roebers et al., 2019).

Metacognition may influence self-derivation performance by changing engagement at any one of the steps leading up to self-derivation. More broadly, metacognition may lead those who are more accurate at monitoring their learning to perform significantly better than their less accurate peers. Even for children who are less successful at self-deriving, accurate metacognitive monitoring can provide knowledge that helps one take steps toward greater success. Because metacognition can inform knowledge about performance and steps one can take to improve performance, a relation reflecting that metacognition facilitates self-derivation may become especially apparent over time. Only a longitudinal investigation provides the opportunity to examine whether metacognition actually drives self-derivation performance. Furthermore, whether this pattern emerges within an individual over development can only be revealed by a longitudinal approach. Because the knowledge that is gained through self-derivation is a product of cognition instead of direct experience, metacognition is a uniquely situated tool to elucidate the age-related and individual variability in self-derivation.

We hypothesized that if a participant is highly accurate at monitoring his performance, then he will also be more accurate at self-deriving new knowledge. If this relation changes across development, then participants will become more accurate at both self-derivation and monitoring their metacognition over the 3 years of the study. We conducted this work with students as they progressed through grades 3-5 for three reasons: 1) children in this age range have previously shown evidence of successful self-derivation across multiple single-sentence trials (Bauer, Blue, Xu, & Esposito, 2017), 2) children in this age range can accurately monitor their learning under some conditions (Metcalfe & Finn, 2013; Roebers & Spiess, 2017; Roebers et al., 2019), and 3)

these children were already participating in a broader longitudinal project including investigation of their self-derivation and academic achievement.

The present research is the first longitudinal examination of the relation between self-derivation performance and metacognition. Children were enrolled at age 8 years and participated each year for 3 years. Children engaged in a single-sentence version of the self-derivation paradigm wherein they encoded individually presented facts. At test, children provided metacognitive judgments. Two metacognitive judgments particularly well-suited to the present investigation are feelings of knowing and confidence judgments. Feelings of knowing (FOK) rely on extrinsic properties of information and allow for understanding how well participants understands both their cognition and the contents of memory. Confidence judgments (CJs) elicited after learning, at test are generally thought to be based on the retrieval experience of testing (Nelson & Narens, 1990; Koriat, 2012), reflecting a veridical measure of participants' metacognitive monitoring. Due to the protracted yet seemingly simultaneous development of both self-derivation and metacognition, it is reasonable to assume that any relation between these processes may also change with age.

Our first aim was to test for replication of previous cross-sectional research (Study 2) investigating the relation between self-derivation and metacognition. The present research was the first opportunity to examine self-derivation performance longitudinally, as well as the first opportunity to examine metacognition for self-derivation longitudinally. Thus, the second aim of this research was to investigate individual consistency in self-derivation, metacognition, and the relation between self-derivation and metacognition during a time of significant development of both these cognitive processes, grades 3 through 5. These aims contribute to the larger goal of

understanding individual variability in self-derivation performance and whether metacognition does indeed facilitate successful self-derivation.

Method

Participants

Participants were 50 students (30 female, 17 male, 3 unreported) in the 3rd grade at the time of initial enrollment. Average age at enrollment was 8.59 years ($SD = .53$). All children were students in the same public school in rural southeastern United States. Caregivers/guardians were provided with informed consent forms through children's parent folders as this is the main form of communication between caregivers and the school system. The caregivers of the children included in the sample provided written informed consent for their children to participate prior to the testing session. Children in the same classroom whose caregivers did not provide consent were tested with their classmates, but their data were not included in analyses. Only children whose caregivers returned a signed consent form are included in analyses. At the end of the session, all children were given a small token of appreciation for their participation (e.g., pencil). All procedures were reviewed and approved by the school board of the participating school system and by the university Institutional Review Board. This process was repeated annually for each of 3 timepoints.

An additional 126 children participated in the study at Year 1 but were not included in the present research because they did not participate and thus contribute data at each of the subsequent waves of data collection. Crucially, there were only 2 significant differences between participants who were included in the present analyses and the participants who were excluded. Participants who were excluded from the present research had significantly higher confidence judgments ($M = 3.09$, $SD = .59$) about their self-derivation performance at Year 1, $t(143) = -2.30$, $p = .02$. Excluded participants were also significantly less accurate in the confidence judgments

made about their stem fact memory at Year 1(M = .27, SD = .61), $t(80)$ = 2.31, p = .02. There were no other significant differences in participants self-derivation or metacognitive measures.

Thirty percent of children's caregivers returned demographic questionnaires on which they reported on the gender and race/ethnicity of their children. Based on parental report, the sample was 28% Black or African American, 32% non-Hispanic Caucasian, and 30% Hispanic or Latinx. Caregivers did not report race and ethnicity for approximately 10% of participants. In the year in which the data were collected, 87% of the children in the participating school system were eligible for free or reduced lunch. Of the families who reported caregiver education, 14% had a high school education or less, 27% had some training beyond high school, 18% had a technical or associates degree, 11% had a college bachelor's degree, and 23% has some training beyond bachelor's degree. Participating parents and teachers were thanked with a $10 gift card to a local merchant.

Stimuli

For each of the three timepoints, stimuli were 16 pairs of true but novel related "stem" facts; unique stimuli were used at each timepoint. Each pair of facts could form an integration fact when integrated. Each pair was about a unique topic, developed based on established curricula and thus appropriate for each grade level. For example, "Venus is the hottest planet" (Stem 1) and "Another name for Venus is morning star" (Stem 2) form an integrable pair. From the integration of these facts, participants can derive the knowledge that "Morning star is another name for the hottest planet." Stimuli changed across grade levels to align with the grade level appropriate curriculum.

Procedure

At each wave, testing took place in children's classrooms over a 45-minute long session. The procedure was administered by two research teams, each consisting of an experimenter and

an assistant. Each research team tested approximately half of all classes. The same two experimenters collected all data over three years. Each subsequent year of testing, one assistant was new; a total of 5 assistants helped to collect data over the 3-year period.

At each wave, the session involved three phases. For a schematic of the experimental session see Figure 1. In the first phase children encoded 16 separate stem facts that were individually projected on a screen at the front of the classroom. Children then completed a 10-minute unrelated buffer activity. In the second phase, children encoded 16 more stem facts and completed a second unrelated buffer activity for 10-minutes.

At the beginning of the third phase, children were instructed that they were going to answer some questions about things they may have learned during the experimenters' visit to the classroom. Children were given a handout of the printed self-derivation questions with spaces for answers and a scale (1-4) for the feeling-of-knowing judgments for each question (Appendix). For Feeling-of-Knowing judgments the Likert-type scale ranged from 1 (definitely will not be able to choose the correct answer in multiple choice) to 4 (definitely will be able to choose the correct answer in multiple choice). A 4-point scale has been used in prior research with children of this age range (Lockl & Schneider, 2002).

The experimenter then oriented participants to the open-ended questions and the metacognitive judgments. First, to demonstrate the low end of the scale, the experimenter provided children with a very difficult question ("When was molybdenum discovered?"). The experimenter explained that because even she, a knowledgeable adult, had no idea what molybdenum is that participants would likely select "1 – not sure at all that I would be able to choose the correct answer in multiple choice." The experimenter then provided an example of an easy question ("What grade are you in?") to prompt children to be very certain they could choose

the correct answer in multiple choice if they saw it later and to demonstrate the high end of the scale. She then explained that she would choose "4 – I would definitely be able to choose the correct answer if the same question was seen again in multiple choice format". Following this orientation, children were instructed to read each question, do their best to provide an answer for each question, and then circle the number corresponding to their feeling-of-knowing. Children self-read the questions and recorded their responses in pencil. The questions were presented in one of four different random orders; children seated next to one another had different versions of the test.

After open-ended testing, the same integration questions were presented in forced-choice format through PowerPoint®. The children were then oriented to the Likert-type scale for the confidence judgments. For confidence judgments the Likert-type scale ranged from 1 (not sure at all that I chose the correct answer) to 4 (completely sure I chose the correct answer).The questions were read aloud by the experimenter and children were asked to select the "best answer" from among three options, one of which was correct (33% expected by chance). Children used individual response devices (a.k.a. "clickers") to provide their responses, which were recorded via TurningPoint® software. The questions were presented in one of four different random orders; each order was used approximately equally often across classrooms. Due to time constraints, we were only able to test recognition of a subset of stem facts (n = 16). Confidence judgments were prompted after each multiple-choice question and children could provide their judgments using the clickers. These procedures were followed at each of the three waves of testing.

Scoring

Children were given 1 point for each self-derivation question answered correctly in open-ended testing. We also calculated scores for children's forced-choice self-derivation performance

and forced-choice stem fact memory with each correct answer yielding 1 point. For each of these three question types, participants answered 16 questions.

Data analysis

From the first timepoint, a subset of self-derivation trials was excluded from all analyses because scoring of participants' open-ended responses revealed that a majority of children provided responses to a given question that could not be scored as evidence for (or the lack of) self-derivation. This frequently was observed as participants providing what would be the correct answer to a stem fact question in place of the correct answer to a self-derivation question. These responses suggest that participants were answering in the right domain but it was unclear whether they had successfully self-derived through integration. We removed 2 trials from Year 1. Because Year 1 included 14 trials and Year 2 and 3 included 16 trials, we converted all self-derivation and stem fact memory scores to proportions.

For each wave, to measure participants' relative metacognitive accuracy, we calculated three gamma correlations for each participant using the Hmisc package in R (Harrell, 2020). Gamma correlations reflect the relative accuracy of metacognitive judgments by comparing pairs of item scores and their judgments based on whether they agree or disagree. For example, a high confidence rating for an item answered correctly would be an agreement whereas a high confidence rating for an item answered incorrectly would be a disagreement. Participants with more agreements than disagreements are more metacognitively accurate than those who have more disagreements or do not vary in their judgments at all. The three gamma correlations calculated measure the accuracy of children's Feeling-of-Knowing judgments provided at open-ended testing and their subsequent forced-choice performance, children's forced-choice self-derivation and their relative confidence in each response, and children's forced-choice stem-fact

memory and their relative confidence in each response. These same three gamma correlations were calculated for each wave of testing.

Each gamma correlation was then correlated to open-ended self-derivation performance, forced-choice self-derivation performance, and forced-choice stem fact memory. The correlation between gamma and performance reflects the strength of the relation between self-derivation or stem fact memory and the accuracy of associated metacognitive judgments.

Multilevel modeling

Multilevel modeling was used to predict individual self-derivation performance from metacognitive accuracy over time. Multilevel modeling is particularly well-suited to the current dataset for at least 3 reasons. First, the dataset is comprised of multiple trials nested within an individual, measured across 3 timepoints. Second, we are interested in predictors at both the level of the trial and at the level of the individual participant; observations within and between years are thus non-independent. Multilevel modeling is well-suited to nested data and violations of the assumption of independent observations (Snijders & Bosker, 1993). MLM is also amenable to missing data whereas many other modeling approaches (e.g., Structural Equation Modeling) are not. Prior research from a similar sample utilized multilevel modeling and models fit appropriately to the data (Bauer, Esposito, & Daly, 2020). Although MLM is ideal, as will become apparent, it was not maximally informative. Thus, we took additional steps and analyzed the data with correlation and regression.

<div align="center">**Results**</div>

The results are presented in 5 parts: 1) change in self-derivation performance and stem fact memory over time, 2) change in average metacognitive judgments and gamma correlations over time, 3) concurrent and cross-lagged relations within self-derivation and metacognitive performance, 4) multilevel models to predict individual self-derivation performance from

metacognitive measures nested within participants over 3 years, and because multilevel modeling was not as informative as expected 5) regression models based on concurrent and cross-lagged relations to predict self-derivation performance from metacognitive measures. All pairwise comparisons are Bonferroni corrected for multiple comparisons.

Self-derivation and stem-fact performance

Self-derivation performance is presented by year in Table 1. Open-ended self-derivation performance was low across all three years; this is in line with prior classroom research (Esposito & Bauer, 2017). Indeed, a repeated measures ANOVA revealed no significant change in self-derivation performance across the 3 years, $F(2, 88) = 1.355$, $p = .263$, $\eta_p^2 = .03$. Because open-ended self-derivation performance is low and reflects little variability, all further analyses utilize only forced-choice data.

Forced-choice self-derivation performance was markedly higher than in open-ended testing. A repeated measures ANOVA showed there was a significant difference in forced-choice self-derivation performance across years, $F(2, 82) = 10.204$, $p < .001$, $\eta_p^2 = .20$. Pairwise comparisons revealed forced-choice performance was significantly higher at Year 2 than at both Year 1 and Year 3. Years 1 and 3 were not significantly different from each other. In spite of changes in forced-choice self-derivation performance over time, performance was significantly correlated. Forced-choice self-derivation performance at Year 1 was related to forced-choice self-derivation performance at Year 2: $r(43) = .54$, $p < .001$ and Year 3: $r(42) = .56$, $p < .001$. Years 2 and 3 are also related with each other, $r(49) = .48$, $p = .001$. Thus participants showed consistency in their relative levels of self-derivation performance over 3 years; high performers tended to successfully self-derive on multiple trials each year while those who did not perform highly at Year 1 tended to remain low-performing.

Stem-fact performance also changed over the three years of the study. However it did so in a pattern that was the inverse of that observed for forced-choice self-derivation: stem-fact memory at Year 2 was significantly lower than at Year 3, $F(2, 56) = 3.383$, $p = .041$, $\eta_p^2 = .11$. Years 1 and 3 were not significantly different from each other. Relative to forced-choice self-derivation performance, stem-fact memory was less consistent over the 3 years of the study. Stem fact memory at Year 1 was not correlated with stem fact performance at Year 2: $r(30) = .08$, $p = .68$ or Year 3: $r(29) = .29$, $p = .13$. However, Years 2 and 3 were significantly correlated with each other, $r(49) = .50$, $p < .001$.

We also examined concurrent and cross-lag relations within the measures of self-derivation task performance. In each year of the study, forced-choice self-derivation and stem fact performance were significantly related: Year 1, $r(30) = .66$, $p < .001$; Year 2: $r(50) = .52$, $p < .001$; and Year 3: $r(49) = .59$, $p < .001$. Each year, the more stem facts participants remembered, the more self-derivation questions they answered correctly. There also were cross-lag relations between the different measures of self-derivation task performance over time. Forced-choice self-derivation at Year 1 was associated with stem fact memory at Year 2: $r(43) = .34$, $p = .02$, and Year 3: $r(42) = .52$, $p < .001$. Self-derivation at Year 2 was significantly related to stem fact memory at Year 3, $r(49) = .65$, $p < .001$.

Metacognitive performance

Children made feeling of knowing (FOK) judgments as predictions about forced-choice self-derivation performance. They also made confidence judgments following forced-choice self-derivation questions (SDCJ) and forced-choice stem fact questions (FCSCJ). Average metacognitive judgments are presented by year in Table 1.

Feeling of Knowing Judgments

Average feeling-of-knowing judgments reflect that children were somewhat certain that they could not accurately predict their performance on the forced choice self-derivation questions (i.e., average of 2 on a 4-point scale). A one-way repeated measures ANOVA showed that there was no change in this predictive judgment over the 3 years, $F(2, 88) = .152$, $p = .859$, $\eta_p^2 = .003$. Feeling of knowing judgments at Year 1 were not related to those at either Year 2: $r(46) = -.10$, $p = .52$ or Year 3: $r(48) = .21$, $p = .15$. However, FOKs at Years 2 and 3 were significantly related, $r(46) = .39$, $p = .008$.

Accuracy of each type of metacognitive judgment is presented as gamma correlations by year in Table 1. At each wave of the study, we calculated gamma correlations for each participant by type of metacognitive judgment. Feeling of knowing judgments serve as a prediction of future selection of the correct answer to forced-choice self-derivation questions. Overall, accuracy of these judgments was low. A one-way repeated measures ANOVA showed there was no significant change in the accuracy of children's feeling-of-knowing judgments over 3 years, $F(2, 52) = .140$, $p = .870$, $\eta_p^2 = .005$. Furthermore, accuracy of FOKs was not related to each other from year to year reflecting inconsistency in the accuracy of children's predictions about their future performance.

Self-derivation Confidence Judgments

On average, children were more confident after they selected their answers for both forced-choice self-derivation and forced-choice stem-fact questions. There was a significant difference in children's average confidence judgments across the three years, $F(2, 64) = 9.12$, $p < .001$, $\eta_p^2 = .22$. Participants were more confident in their self-derivation performance at Year 1 and Year 2 compared to Year 3. There was no difference in confidence between Years 1 and 2. Overall, confidence following self-derivation questions decreased over the duration of the study.

In spite of the general negative trend, confidence judgments about self-derivation performance at Year 1 were related with those at Year 2: $r(34) = .40$, $p = .02$ and Year 3: $r(37) = .36$, $p = .03$. Self-derivation confidence judgments were not related between Years 2 and 3, $r(42) = .25$, $p = .12$, however.

Despite the change over time in average confidence judgment following forced-choice self-derivation questions, we did not observe a significant change in the accuracy of these judgments, $F(2, 58) = 1.758$, $p = .181$, $\eta_p^2 = .06$. We did observe a significant difference in the accuracy of children's confidence in their forced-choice stem-fact performance $F(2, 16) = 6.004$, $p = .011$, $\eta_p^2 = .43$. Children were significantly more accurate in Year 1 compared to Year 2. There was no difference between Years 2 and 3 or between Years 1 and 3. Children did not exhibit age-related increases in the accuracy of their metacognitive judgments. Just as the accuracy of feeling of knowing judgments was not related from year to year, neither was the accuracy of self-derivation confidence judgments or stem fact confidence judgments. This provides strong evidence that children's metacognition is not stable and is continuing to develop beyond age 11 years.

Stem Fact Memory Confidence Judgments

We did not observe a significant effect of time on children's average confidence judgment following forced-choice stem fact questions, $F(2, 24) = 2.748$, $p = .084$, $\eta_p^2 = .19$. Notably, confidence judgments elicited after stem fact questions were most likely to be skipped by children leading to a smaller sample of these judgments and less power to detect an effect. Confidence judgments about stem fact memory at Year 1 were not correlated with those at either Year 2, $(15) = .40$, $p = .14$, or Year 3, $r(37) = .20$, $p = .24$.

At Year 1, confidence judgments about self-derivation were correlated with confidence judgments about stem fact memory, $r(17) = .77$, $p < .001$. They were not related at Year 2: $r(42) = .04$, $p = .81$. At Year 3 confidence judgments about self-derivation and stem fact performance were again correlated, $r(49) = .62$, $p < .001$.

Correlation matrices for each year are provided in Tables 2-4. Table 5 represents cross-year correlations.

Year 1 feeling of knowing judgments were not significantly correlated with confidence judgments about self-derivation performance, $r(37) = .31$, $p = .06$, or about stem fact memory, $r(17) = .04$, $p = .87$. Year 2 FOKs were significantly related to confidence judgments about self-derivation performance, $r(40) = .38$, $p = .02$; Year 2 FOKs were not related to confidence judgments about stem fact memory, $r(39) = .07$, $p = .67$. At Year 3 FOKs were related to confidence judgments about self-derivation performance, $r(48) = .39$, $p = .006$, and about stem fact memory, $r(48) = .39$, $p = .007$.

Year 1 confidence judgments about self-derivation were not correlated with confidence judgments about stem fact memory at Year 2: $r(33) = -.16$, $p = .39$ or Year 3: $r(37) = .20$, $p = .24$. Similarly, Year 2 confidence judgments about self-derivation performance were not correlated with confidence in stem fact memory at Year 3, $r(42) = .25$, $p = .11$.
Feeling of knowing judgments at Year 1 were significantly related to confidence judgments about self-derivation performance at Year 2, $r(42) = .38$, $p = .01$. They were also related to self-derivation confidence judgments at Year 3, $r(48) = .41$, $p = .004$. Confidence judgments about stem fact memory at Year 1 were not related to feeling of knowing judgments at Year 2: $r(17) = .15$, $p = .56$, or at Year 3: $r(16) = -.11$, $p = .67$. These findings suggest there is limited stability in metacognition over the time between 3rd and 5th grades.

Relations between Metacognition and Self-derivation

At Year 1, forced-choice self-derivation was significantly related to accuracy of the self-derivation confidence judgments. This relation was not significant at Year 2. At Year 3, forced-choice self-derivation performance was again related to the accuracy of post self-derivation confidence judgments $r(49) = .29, p < .05$.

Accuracy of confidence judgments about self-derivation performance at Year 1 is significantly related to forced-choice self-derivation performance and stem fact memory at Year 3. Accuracy of stem fact memory confidence judgments at Year 2 is significantly related to accuracy of confidence judgments following self-derivation questions at Year 3.

Multilevel Models

To predict self-derivation performance over time with metacognition and to account for the nested structure of trials within participants collected at multiple timepoints, we used multilevel modeling. We first tested a null model to determine whether there was sufficient variance at both levels of the model to predict self-derivation performance. First, we constructed a null model with individual forced-choice self-derivation trials at Level 1 and participant at Level 2 to calculate the variance between participants. The intraclass correlation for this null model reflected 8% of the variance in self-derivation performance over all observations was between participants. Approximately 92% of the variance remained to be explained by variables at Level 1, the level of the individual trial. We added Year to the model as a Level 1 predictor to account for the change within participants over time. Because the addition of Year as a predictor reduced the variance between participants, we then added Year in as a random slope. This allows for individual participants' self-derivation to vary across years. When metacognitive variables were entered into this model, the model failed to converge due to overparameterization. The null model is presented in Table 6.

Because Year was a significant predictor but we could not evaluate multiple years in the same model, we then specified 3 separate models predicting self-derivation from metacognitive accuracy for each year. Models for each year of data collection are presented in Table 7. Both the accuracy of feeling of knowing judgments and of confidence judgments were significant predictors of self-derivation performance within each year's model. This is demonstrated by lower AIC, BIC, and -2LL, displayed in Table 7, when the predictors are added to the models.

Ultimately, these models suggest that metacognitive accuracy does explain some of the variance in self-derivation performance but because we could not construct a model containing data from all 3 years of data collection, our next step was to use regression to investigate self-derivation and metacognition over time.

Regression Models

Because multilevel modeling did not allow for successful examination of change in the relation between self-derivation and metacognition over time, we then conducted 3 multiple regression analyses. Full regression models are provided in Tables 8 through 10.

In the first regression model, we predicted Year 2 forced-choice self-derivation performance from measures at Year 1. The overall model significantly predicted self-derivation performance at Year 2, $F(2, 29) = 4.12$, $p = .03$, $R^2 = .23$. Predictors entered were Year 1 forced-choice self-derivation performance ($B = .40$, $SE = .18$, $p = .04$) and stem fact performance ($B = .03$, $SE = .12$, $p = .81$). Of these measures, only self-derivation performance at Year 1 was a significant predictor of Year 2 self-derivation performance. The full model is reported in Table 8.

In the second model, we predicted Year 3 self-derivation from measures at Year 1. The predictors were Year 1 forced-choice self-derivation and FCICJ gamma. The model fit well, $F(2, 33) = 13.27$, $p < .001$, $R^2 = .46$. Forced choice self-derivation ($B = .49$, $SE = .13$, $p = .001$) was

the only significant predictor. The full model is reported in Table 9. In the third model, we predicted Year 3 self-derivation from measures at Year 2. The predictors were forced-choice self-derivation and stem fact memory. The model was significant, $F(2, 48) = 7.57$, $p = .001$, $R^2 = .248$. Only forced-choice self-derivation performance ($B = .33$, $SE = .13$, $p = .01$) was a significant predictor. The full model is reported in Table 10.

Discussion

The present research was a longitudinal investigation of self-derivation and metacognition in elementary school students. This study had two primary aims. The first was to examine the association between self-derivation and metacognition, effectively testing for replication of prior research. The second aim was to examine consistency in self-derivation, metacognition, and the relation between self-derivation and metacognition within individuals over time. This approach allows systematic investigation of the individual differences in self-derivation and whether metacognition facilitates self-derivation performance. This study expands our understanding of self-derivation and metacognition as it is the first longitudinal investigation of self-derivation as well as the first longitudinal examination of metacognition about self-derivation.

We first examined change in self-derivation performance across the three years of the study. Forced-choice self-derivation performance was significantly higher at Year 2 than at Years 1 or 3. Based on prior research reflecting age-related change in self-derivation (Bauer & Larkina, 2017; Esposito & Bauer, 2017), the increase in performance from Year 1 to Year 2 is as we expected. However, self-derivation at Year 3 was not different from Year 1. It is likely that this lower performance at Year 3 is due to changes in the stimuli. For each year of testing, the stimuli were aligned to the grade-level curriculum. Because we could not use the same stimuli

every year and the difficulty of the information learned was appropriate for the grade-level, it may be that the material learned at Year 3 was sufficiently challenging that we could not observe age-related change in self-derivation performance.

Interestingly, we observed a different pattern in participants' stem fact memory. At Year 1, participants successfully remember stem facts but exhibit low levels of self-derivation. The following year children integrated the stem facts to self-derive new knowledge but did not remember the individual stem facts. Despite experiencing the highest levels of self-derivation performance at Year 2, stem fact memory was then at its lowest. This suggests that perhaps children remembered the facts necessary to self-derive in Years 1 and 3 but they did not go on to integrate them and successfully generate new knowledge. Stem fact memory was not correlated with forced-choice self-derivation at Year 1 but it was at both Year 2 and Year 3. Again in Year 3, children remembered the stem facts but they did not self-derive, perhaps due to the difficult nature of the curriculum-aligned stimuli. Together, these data suggest age-related changes in engagement of the subprocesses of self-derivation.

The longitudinal design of the current study afforded new insight into self-derivation across development. Within participants we observed that stem fact memory was not correlated between Years 1 and 2 but it was correlated between Years 2 and 3. The emergence of a correlation in stem fact memory over time suggests increasing consistency in remembering stem facts, one of the necessary components of successful self-derivation.

One of the most striking results of the present research is the robust stability in self-derivation performance from year to year. Correlations demonstrated that forced-choice self-derivation performance at Year 1 predicts self-derivation at both Year 2 and Year 3. Indeed, in regression models, self-derivation performance at Year 1 was the only significant predictor of

subsequent self-derivation in all of the models we analyzed. Stability in the individual differences of self-derivation is evident in the present sample. Perhaps most surprising about this pattern of findings is that stability in self-derivation was apparent prior to stability in memory for separate but related stem facts. Because stem facts are necessary but not sufficient for the generation of new factual knowledge, one might expect that stem fact memory would be consistent over time before the integration of these stem facts stabilized. We had fewer stem fact observations than self-derivation observations; it is possible that with more stem fact observations, stability in stem fact memory would be apparent.

The present research also served as the first longitudinal investigation of metacognition about self-derivation. The first step in elucidating metacognition for self-derivation was an examination of children's metacognitive judgments. Overall, children's predictions of how well they would be able to select the correct answer to a forced-choice self-derivation question were low. They most often chose "I would probably not be able to select the correct answer." Their confidence in their learning rose after the open-ended test for self-derivation. After forced-choice testing the average judgment was "I'm pretty sure I chose the correct answer." This pattern held across Year 1 and Year 2. By Year 3, children were significantly less confident in their forced-choice performance. This overall decrease in confidence is not surprising given prior research suggesting that children become less overconfident about their performance over development (Roebers & Spiess, 2017). Unlike confidence in self-derivation performance, there was no difference in stem fact memory confidence over time. This suggests that over time children may grow in their understanding of self-deriving and perhaps they gain some sense that generating new knowledge is more challenging than learning individual facts directly.

When examining the accuracy of children's metacognitive judgments, we found no change over time in either the accuracy of feeling of knowing or of self-derivation confidence judgments. This underscores both the challenging nature of self-deriving to gain new knowledge and the challenge of monitoring one's self-derivation as it is distributed over time. Self-derivation requires integrating facts learned at separate timepoints. The new knowledge that results from successful self-derivation is a product of one's mind rather than direct experiences alone suggesting that in order to successfully monitor this process, one must be acutely aware of her own cognition for an extended duration. In regards to stem fact performance, we did observe an age-related change in the accuracy of children's confidence judgments; however, children appeared to become less accurate over time. At Years 1 and 3, children had an average gamma correlation of .68 and .50, respectively. Notably, in studies of children's metacognition, participants typically engage with word pairs, pictures, spelling, facts, and learning definitions – tasks that are typically far less challenging than creating entirely new knowledge.

The longitudinal design of the current study afforded examination of the intraindividual consistency in metacognition about self-derivation. Children's confidence in self-derivation and stem fact performance were related in Year 1, but neither was related to the feeling of knowing predictions children made during open-ended testing for self-derivation. By Year 2, the pattern shifted such that children's confidence in their self-derivation was no longer related to confidence in their stem fact memory but was related to feelings of knowing judgments. It is possible that children relied on different information provided by metacognitive monitoring across time. In the final year of data collection, confidence in self-derivation performance was related to both feelings of knowing and stem fact confidence judgments. Unlike the patterns observed in the relations within self-derivation measures, metacognition did not follow a

consistent pattern of relations each year. This suggests that there may be meaningful change in metacognition as it relates to self-derivation over the 3-year period in which we observed children. Unlike self-derivation, metacognition as we measured it was not stable within individual participants across the duration of the study; metacognition at earlier timepoints did not relate to later metacognition. More specifically, the only significant association we observed in metacognitive accuracy was between accuracy of stem fact confidence judgments at Year 2 and accuracy of confidence in self-derivation at Year 3. Recent results from a cross-sectional and longitudinal study by Roebers et al., (2019) suggest that the lack of stability in metacognition in the present research is unsurprising. Roebers et al., (2019) measured both 2nd graders' and 4th graders' metacognitive monitoring of their learning novel Japanese characters; monitoring at early timepoints did not predict later monitoring, even in the 4th graders.

Upon inspection of the relation between metacognitive accuracy and self-derivation performance, we found that it was only for Year 1 and Year 3 that participants' accuracy of their self-derivation confidence judgments was significantly related to their concurrent forced-choice self-derivation performance. Moreover, across timepoints, the only significant relation we observed was between accuracy of confidence in self-derivation performance at Year 1 and self-derivation performance at Year 3. In other words, there was not a pattern of between-task relations such that self-derivation predicted metacognition or that metacognition predicted self-derivation within or across timepoints.

As previously discussed, there was strong evidence of stability in children's self-derivation performance but not in their metacognition. Children's self-derivation at Year 1 predicted performance at both Year 2 and Year 3. In contrast, we did not observe stability within metacognition. Metacognition at Year 1 and Year 2 did not predict subsequent metacognitive

performance. Neither self-derivation performance nor accuracy of confidence in self-derivation performance changed significantly over time but the relation between these two variables changed from year to year such that self-derivation performance was related to concurrent accuracy of confidence in self-derivation performance at Years 1 and 2 but not at Year 3. It is unclear what might account for this coupling and decoupling. However, decoupling of the relation between self-derivation performance and metacognitive accuracy at Year 2 suggests that for children in the observed age range, self-derivation is not driven by metacognition. It is possible that if metacognition develops more slowly than self-derivation, as suggested by the lack of stability in metacognition over time, then children are not yet able to utilize metacognition to direct successful self-derivation performance. processes. We further postulate that this pattern of coupling and decoupling may be due to changes in other cognitive processes relevant for success in both self-derivation and metacognition, such as working memory (Klesczewski et al., 2018). Extending the present research further into adolescence will help shed light on the factors contributing to the changes in the relation between self-derivation performance and metacognitive accuracy over time.

The low accuracy of feeling of knowing judgments and the lack of relation between accuracy and self-derivation performance reflects that children between 8- and 11-years-old are more adept at judging their performance after it has occurred. Taken together, these findings suggest that although there are established patterns of development of metacognitive monitoring in the laboratory during the elementary school years and robust consistency in self-derivation over time metacognitive abilities are not yet stable enough within individuals to make clear contributions to self-derivation performance.

In summary, we conducted a longitudinal investigation of the relation between self-derivation and metacognition. We observed an increase in self-derivation performance from Year 1 to Year 2 but not from Year 2 to Year 3. Despite the drop in performance at Year 3, we

observed intraindividual stability in self-derivation performance. This stability was not observed in metacognition across all 3 years of the study. At present there is limited data demonstrating a significant relation between self-derivation and metacognition either concurrently or over time. We hypothesized that metacognition would facilitate self-derivation success and that the relation between metacognition and self-derivation that allowed for such facilitation would increase over time. As evidenced by the relation between self-derivation performance and confidence judgment accuracy at Years 1 and 3, the present results suggest that metacognition can contribute to the individual variability observed in self-derivation. However, during this period of development, metacognition is not yet a significant contributor to successful self-derivation performance. Lack of consistency in metacognition over the observed 3-year period makes it difficult to ascertain just how metacognition contributes to self-derivation. Two points make clear that this line of research is worth further investigation. Both self-derivation and metacognition continue developing beyond 5th grade, the latest timepoint in the present research, providing future opportunities for a relation to develop. Furthermore, self-derivation is a unique process by which individuals author their own learning and knowledge creation and metacognition is a tool uniquely suited to provide insight into the experience and understanding of this process.

Limitations and Future Directions

One frequent limitation of longitudinal studies is missing data (Enders, 2012; Jeličić, Phelps, & Lerner, 2009; Rovine & Delaney, 1990). Given the nature of the metacognitive data, we did not find techniques to correct for missing data to be appropriate. These techniques often use estimation based on a participant's average response to an item; because we were interested in the variation in metacognitive judgments between items, such techniques may have biased a

participants' metacognitive accuracy and rendered the metacognitive measures less informative. Approximately 180 children were tested during Year 1 of the present research, but due to children transferring schools, being absent, or not having parental consent, our final sample was only 50 children on whom we had data for all three waves of the study. Crucially, participants in the present study did not differ from those who were excluded in meaningful ways. It was only confidence in self-derivation and accuracy of stem fact confidence that differed between these two groups; these differences may have been driven by the unequal sample sizes of the two groups.

The present research is part of a larger longitudinal study on children's learning. Metacognitive judgments in the current study were elicited only at test for self-derivation; we did not have children provide judgments during the learning phase. We were constrained to the test phase of the study to prevent any impact on encoding of the stem facts and subsequent test for self-derivation. Metacognition may influence self-derivation at any one of several steps to self-derivation. For this reason, we must systematically examine metacognition at the observable steps of self-derivation. Such an approach will lend clarity to our understanding of how metacognition supports self-derivation performance across development.

Future research will aim to further elucidate children's metacognition about self-derivation, as well as their overall metacognitive abilities, to create a broader understanding of how the two interact. Furthermore, future research should investigate metacognition as a potential mechanism to target for improved knowledge acquisition. Because metacognition is developing in children during grades 3 through 5, providing them with a lesson on engaging in monitoring may yield clearer evidence of the association between self-derivation and metacognition. Extending the duration of the study to span grades K through 7 may also provide

the context in which to observe minute changes in metacognition and how it may relate to self-derivation performance.

Conclusion

In conclusion, the present research provided the first step toward understanding the relation between self-derivation and metacognition in individual children, across development. Self-derivation performance increased initially but did not continue increasing with time. Self-derivation performance was also consistent within individuals. These results stand in contrast to our metacognitive findings. There was limited evidence of change in metacognition over development or of a robust relation to self-derivation. Future research must be conducted to thoroughly explicate the relation between self-derivation and metacognition, but the present research provides the foundation on which to expand.

References

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

Bauer, P. J., Esposito, A. G., & Daly, J. J. (2020). Self-derivation through memory integration: A model for accumulation of semantic knowledge. *Learning and Instruction*, *66*, 101271.

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. doi:10.1037/a0037405

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

Bauer, P. J., & San Souci, P. (2010). Going beyond the facts: Young children extend knowledge by integrating episodes. *Journal of Experimental Child Psychology*, *107*(4), 452-465.

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*. doi:10.1016/j.jarmac.2017.01.006

Bauer, P. J., Cronin-Golomb, L. M., Porter, B. M., Jaganjac, A., & Miller, H. E. (in press) Developmental differences in the conditions of memory integration in adults and children: Evidence across indirect and direct measures. *Journal of Experimental Psychology: General*

Dugan, J. A., & Bauer, P. J. (2020) The process of self-derivation: Insights from 8-year-olds' eye movements, memory, and source attributions. *Manuscript in preparation.*

Enders, C. K. (2013). Dealing with missing data in developmental research. *Child Development Perspectives*, *7*(1), 27-31.

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.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*(10), 906.

Goswami, U. (2011). Inductive and deductive reasoning. In U. Goswami (Ed.), Childhood cognitive development (pp. 399–419). Oxford, UK: Wiley–Blackwell. doi.org/10.1016/j.tics.2010.10.

Harrell, F. E.(2020). Hmisc: Harrell Miscellaneous. R package version 4.4-0. https://CRAN.R-project.org/package=Hmisc

Jeličić, H., Phelps, E., & Lerner, R. M. (2009). Use of missing data methods in longitudinal studies: the persistence of bad practices in developmental psychology. *Developmental Psychology*, *45*(4), 1195.

Klesczewski, J., Brandenburg, J., Fischbach, A., Schuchardt, K., Grube, D., Hasselhorn, M., & Büttner, G. (2018). Development of working memory from grade 3 to 5: Differences between children with and without mathematical learning difficulties. *International Journal of Disability, Development and Education*, *65*(5), 509-525.

Koriat, A. (2012). The subjective confidence in one's knowledge and judgments: some metatheoretical considerations. In M. Beran, J. L. Brandl, J. Perner, & J. Proust (Eds.), *The Foundations of Metacognition* (pp. 213-233). Oxford: Oxford University Press.

Metcalfe, J. & Finn, B. (2013). Metacognition and control of study choice in children. *Metacognition and Learning, 8*(1), 19-46.

Nelson, T.O. & Narens, L. (1990). Metamemory: A theoretical framework and some new findings. In G.H. Bower (Ed). *The Psychology of Learning and Motivation, 26*, 125-173. New York: Academic Press

Roebers, C. M. & Spiess, M. (2017). The development of metacognitive monitoring and control in second graders: A short-term longitudinal study. *Journal of Cognition and Development, 18*, 110-128.

Roebers, C. M., Mayer, B., Steiner, M., Bayard, N., & Van Loon, M. (2019). The role of children`s metacognitive experiences for cue utilization and monitoring accuracy: A longitudinal study. *Developmental Psychology. 55*(10), 2077-2089.

Rovine M. & Delaney M. (1990). Missing data estimation in developmental research. In: Alexander von Eye (Ed.). *Statistical Methods in Longitudinal Research* (Vol. 1). Academic Press; Boston.

Schneider, W., & Lockl, K. (2002). *The development of metacognitive knowledge in children and adolescents.* In T. J. Perfect & B. L. Schwartz (Eds.), *Applied Metacognition* (p. 224–257). Cambridge University Press. https://doi.org/10.1017/CBO9780511489976.011

Schneider, W. (1985). Developmental trends in the metamemory-memory behavior relationship: An integrative review. In D. Forrest-Pressley, G. E. MacKinnon, & T. G. Waller (Eds.), *Metacognition, Cognition, and Human Performance Vol 1* (p. 57-109). Academic Press; Boston.

Snijders, T. A., & Bosker, R. J. (1993). Standard errors and sample sizes for two-level research. *Journal of Educational Statistics*, *18*(3), 237-259.

Tsalas, N., Paulus, M., & Sodian, B. (2015). Developmental changes and the effect of self-generated feedback in metacognitive controlled spacing strategies in 7-year-olds, 10-year-olds, and adults. *Journal of Experimental Child Psychology*, *132*, 140-154.

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. (2013). Effects of delays on 6-year-old children's self-generation and retention of knowledge through integration. *Journal of Experimental Child Psychology*, *115*(2), 326-341.

Varga, N. L., Esposito, A. G., & Bauer, P. J. (2020). Cognitive correlates of memory integration across development: Explaining variability in an educationally relevant phenomenon. *Journal of Experimental Psychology: General*, *148*(4), 739.

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.

Table 1

Self-derivation and Metacognition Measure Descriptives

Measure	Year 1	Year 2	Year 3
OE Self-derivation Proportion Correct	.10 (.11)	.12 (.13)	.10 (.14)
FC Self-derivation Proportion Correct	.46 (.18)	.58 (.18)	.49 (.16)
FC Stem fact memory Proportion Correct	.57 (.23)	.47 (.26)	.59 (.21)
Mean FOK	2.11 (.59)	2.11 (.60)	2.09 (.63)
Mean FCICJ	2.83 (.70)	2.69 (.56)	2.32 (.57)
Mean FCSCJ	3.21 (.72)	2.79 (.87)	2.75 (.56)
Mean FOK accuracy	.16 (.55)	.18 (.57)	.33 (.58)
Mean FCSD-CJ accuracy	.24 (.54	.41 (.52)	.27 (.49)
Mean FCS-CJ accuracy	.67 (.34)	-.01 (.56)	.50 (.57)

Table 2

Year 1 Correlations

Panel A

	OE Self-derivation	FC Self-derivation	FC Stem Fact memory
OE Self-derivation	1	.601**	0.341
FC Self-derivation		1	.584**
FC Stem Fact Memory			1

Panel B

	FOK	FCSD-CJ	FCS-CJ	FOK Accuracy	FCSD-CJ Accuracy	FCS-CJ Accuracy
Feeling of Knowing	1	.309	.042	.108	-.138	.155
FC Self-derivation CJ		1	.766**	.103	-.100	.136
FC Stem Fact Memory CJ			1	-.041	-.013	.218
FOK Accuracy				1	.147	-.127
FCSD-CJ Accuracy					1	.084
FCS-CJ Accuracy						1

Panel C

	OE Self-derivation	FC Self-derivation	FC Stem Fact Memory	FOK Accuracy	FCSD-CJ Accuracy	FCS-CJ Accuracy
OE Self-derivation	1	.601**	0.341	0.024	.511**	-0.11
FC Self-derivation		1	.584**	-0.133	.341*	0.39
FC Stem Fact Memory			1	-0.094	0.22	0.241
FOK Accuracy				1	0.147	-0.127
FCSD-CJ Accuracy					1	0.084
FCS-CJ Accuracy						1

Note. **. Correlation is significant at the 0.01 level (2-tailed). *. Correlation is significant at the 0.05 level (2-tailed).

Table 3

Year 2 Correlations

Panel A

	OE Self-derivation	FC Self-derivation	FC Stem Fact Memory
OE Self-derivation	1	.683**	.556**
FC Self-derivation		1	.517**
FC Stem Fact Memory			1

Panel B

	FOK	FCSD-CJ	FCS-CJ	FOK Accuracy	FCSD-CJ Accuracy	FCS-CJ Accuracy
Feeling of Knowing	1	.380*	0.071	0.094	-0.11	0.081
FC Self-derivation CJ		1	0.038	.359*	0.217	0.131
FC Stem Fact Memory CJ			1	0.102	-0.002	0.144
FOK-FCSD Accuracy				1	0.115	0.346
FCSD-CJ Accuracy					1	0.216
FCS-CJ Accuracy						1

Panel C

	OE Self-derivation	FC Self-derivation	FC Stem Fact Memory	FOK Accuracy	FCSD-CJ Accuracy	FCS-CJ Accuracy
OE Self-derivation	1	.683**	.556**	0.306	0.216	.464**
FC Self-derivation		1	.517**	0.284	0.261	0.303
FC Stem Fact Memory			1	0.082	0.25	0.183
FOK-FCSD Accuracy				1	0.115	0.346
FCSD-CJ Accuracy					1	0.216
FCS-CJ Accuracy						1

Note. **. Correlation is significant at the 0.01 level (2-tailed). *. Correlation is significant at the 0.05 level (2-tailed).

Table 4

Year 3 Correlations

Panel A

	OE Self-derivation	FC Self-derivation	FC Stem Fact Memory
OE Self-derivation	1	.667**	.614**
FC Self-derivation		1	.590**
FC Stem Fact Memory			1

Panel B

	FOK	FCSDCJ	FCSCJ	FOK Accuracy	FCSD-CJ Accuracy	FCS-CJ Accuracy
Feeling of Knowing	1	.391**	.387**	0.091	.342*	0.195
FC Self-derivation CJ		1	.621**	-0.091	0.21	0.087
FC Stem Fact Memory CJ			1	-0.064	0.228	-0.182
FOK-FCSD Accuracy				1	0.159	0.173
FCSD-CJ Accuracy					1	-0.052
FCS-CJ Accuracy						1

Panel C

	OE Self-derivation	FC Self-derivation	FC Stem Fact Memory	FOK Accuracy	FCSD-CJ Accuracy	FCS-CJ Accuracy
OE Self-derivation	1	.667**	.614**	0.255	0.257	-0.1
FC Self-derivation		1	.590**	0.234	.292*	0.175
FC Stem Fact Memory			1	0.125	0.254	-0.016
FOK Accuracy				1	0.159	0.173
FCSD-CJ Accuracy					1	-0.052
FCS-CJ Accuracy						1

Note. **. Correlation is significant at the 0.01 level (2-tailed). *. Correlation is significant at the 0.05 level (2-tailed).

Table 5

Cross-lagged Correlations between Self-derivation and Metacognitive Measures

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
OE SD Year 1	1	.601**	0.341	.501**	.564**	.454**	.624**	.535**	.581**	0.024	.511**	-0.11	-0.112	0.194	0.081	0.165	0.177	0.07
FC SD Year 1		1	.584**	.443**	.541**	.344*	.613**	.561**	.518**	-0.133	.341*	0.39	-0.145	0.105	0.233	0.321	0.088	0.166
FC Stem Year 1			1	0.093	0.315	0.079	0.362	0.192	0.292	-0.094	0.22	0.241	-.480*	0.258	0.138	.589**	0.044	0.02
OE SD Year 2				1	.683**	.556**	.565**	.379**	.530**	0.176	0.089	-0.332	0.306	0.216	.464**	.331*	.294*	-0.25
FC SD Year 2					1	.517**	.617**	.475**	.650**	-0.123	0.221	0.226	0.284	0.261	0.303	0.171	0.245	-0.261
FC Stem Year 2						1	.441**	.374**	.495**	0.068	0.227	-0.345	0.082	0.25	0.183	0.19	0.104	-0.09
OE SD Year 3							1	.667**	.614**	0.031	.367*	-0.115	-0.032	0.122	0.124	0.255	0.257	-0.1
FC SD Year 3								1	.590**	0.005	.463**	-0.105	-0.157	0.026	0.224	0.234	.292*	0.175
FC Stem Year 3									1	-0.092	.483**	-0.082	0.095	0.139	0.171	0.125	0.254	-0.016
FOK Acc Y1										1	0.147	-0.127	-0.125	0.184	-0.094	0.016	-0.055	-0.262
FCSD-CJ Acc Y1											1	0.084	-0.077	0.319	0.041	0.001	0.074	-0.01
FCS-CJ Acc Y1												1	-0.312	0.057	0.091	-0.231	-0.241	0.206
FOK Acc Y2													1	0.115	0.346	0.193	0.094	-0.145
FCSD-CJ Acc Y2														1	0.216	0.091	0.042	-0.184
FCS-CJ Acc Y2															1	0.279	.412*	0.02
FOK Acc Y3																1	0.159	0.173
FCSD-CJ Acc Y3																	1	-0.052
FCS-CJ Acc Y3																		1

Table 6

Null Model of Self-derivation Performance by Grade

Overall null model

	τ_{00}	σ^2	AIC	BIC	-2LL
Null Model	8%	92%	5571.3	5587.7	5565.3

	γ_{10}	t	p	AIC	BIC	-2LL
Grade as Intercept	.84	28.61	< .001	4902.1	4923.9	4894.1
Grade as Slope	.07	9.457	<.001	4854.7	4887.5	4842.7

Level 1: Self-derivation$_{it}$ = β_{0it} + r$_{it}$

Level 2: β_{0i} = γ_{00} + u$_{0i}$

B$_{0i}$ is the mean self-derivation for participant i, γ_{00} is the overall mean self-derivation of the sample, and u$_{0i}$ represented the degree to which individuals vary from the sample as a whole.

Table 7

Multilevel Models of Self-derivation Performance as Predicted by Metacognitive Measures Each Year

Panel A Model by Year

	τ_{00}	σ^2	AIC	BIC	-2LL
Year					
1	6%	94%	864.9	878.1	858.9
2	8%	92%	1124.9	1139.0	1118.9
3	4%	96%	1136.1	1150.1	1130.1

Panel B Feeling of Knowing as a single predictor

	γ_{10}	t	p	AIC	BIC	-2LL
Year						
1	.05	7.192	< .001	671.9	688.5	663.9
2	.07	9.457	<.001	808.9	826.4	800.9
3	.10	4.769	<.001	782.9	800.2	774.9

Panel C Feelings of Knowing and Confidence Judgments as Predictors of FC Self-derivation

	γ_{10}	t	p	AIC	BIC	-2LL
Year						
1 – FOK	.05	2.152	< .05	509.7	529.2	499.7
1 – CJ	.10	4.300	<.001			
2 – FOK	.05	2.298	<.05	559.3	579.7	549.3
2 – CJ	.15	7.843	<.001			
3 – FOK	.07	3.170	<.01	698.8	719.9	688.8
3 – CJ	.08	3.739	<.001			

Model 1 The random coefficients regression models equations used to test whether metacognition predicted self-derivation were:

Level 1: Self-derivation$_{it}$ = β_{0it} + β_{1it}(metacognition) + r_{it}

Level 2: β_{0i} = γ_{00} + u_{0i}

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

The slope coefficient β_{1it} represented the associated change in self-derivation associated with change in metacognitive 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 self-derivation for the sample. Further, γ_{10} corresponded to the effect of metacognition on self-derivation and u_{0i} and u_{1i} represent the degree to which individuals vary from the sample as a whole.

Table 8

Regression Analysis Summary for Year 1 Self-derivation and Metacognition Predicting Self-derivation at Year 2

Variable	B	95% CI	β	t	p
Constant	0.40	[0.25, 0.55]		5.37	0.000
Year 1 FC Self-derivation	0.40	[0.02, 0.77]	0.45	2.18	0.04
Year 1 Stem Fact Memory	0.29	[-0.21, 0.27]	0.05	0.25	.81

Note. $R^2_{adjusted} = .177$, CI = Confidence Interval for B

Table 9

Regression Analysis Summary for Year 1 Self-derivation and Metacognition Predicting Self-derivation at Year 3

Variable	B	95% CI	β	t	p
Constant	0.23	[0.10, 0.35]		3.54	0.001
Year 1 FC Self-derivation	0.49	[0.23, 0.76]	0.53	3.77	0.001
Year 1 FCSD-CJ Gamma	0.09	[-0.001, 0.19]	0.28	2.03	0.05

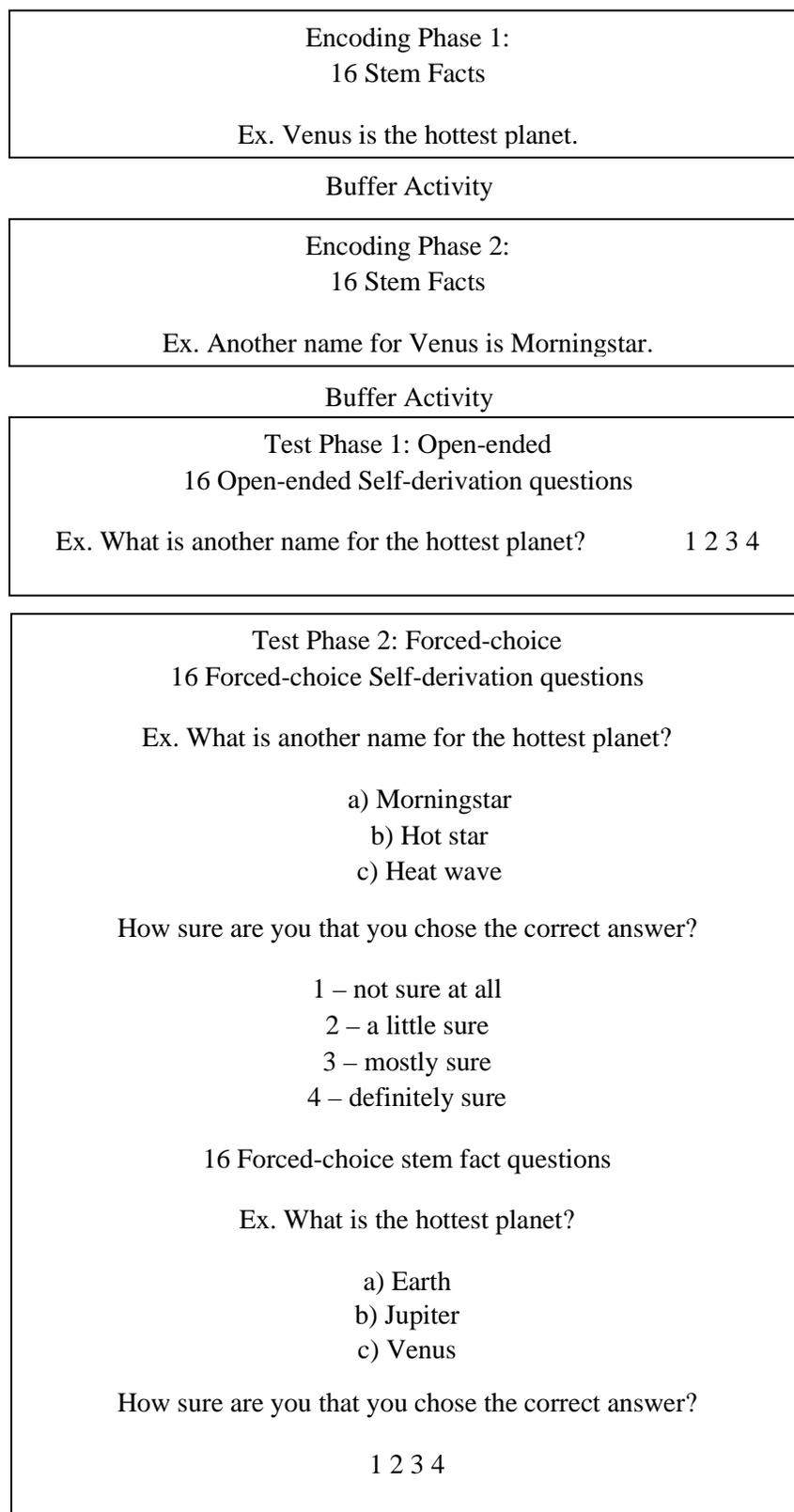*Note.* $R^2_{adjusted} = .427$, CI = Confidence Interval for B

Table 10

Regression Analysis Summary for Year 2 Self-derivation and Metacognition Predicting Self-derivation at Year 3

Variable	B	95% CI	β	t	p
Constant	0.24	[0.11, 0.38]		3.63	0.001
Year 2 FC Self-derivation	0.33	[0.07, 0.59]	0.53	2.57	0.01
Year 2 Stem Fact Memory	0.11	[-0.08, 0.30]	0.28	1.17	0.25

Note. $R^2_{adjusted} = .215$, CI = Confidence Interval for B

Figure 1. Schematic of Procedure in Classroom

| Encoding Phase 1: |
| 16 Stem Facts |
| |
| Ex. Venus is the hottest planet. |

Buffer Activity

| Encoding Phase 2: |
| 16 Stem Facts |
| |
| Ex. Another name for Venus is Morningstar. |

Buffer Activity

| Test Phase 1: Open-ended |
| 16 Open-ended Self-derivation questions |
| |
| Ex. What is another name for the hottest planet?   1 2 3 4 |

Test Phase 2: Forced-choice
16 Forced-choice Self-derivation questions

Ex. What is another name for the hottest planet?

a) Morningstar
b) Hot star
c) Heat wave

How sure are you that you chose the correct answer?

1 – not sure at all
2 – a little sure
3 – mostly sure
4 – definitely sure

16 Forced-choice stem fact questions

Ex. What is the hottest planet?

a) Earth
b) Jupiter
c) Venus

How sure are you that you chose the correct answer?

1 2 3 4

General Discussion

The present research was the first series of direct examinations of the relation between self-derivation and metacognition across development. The aims of this research were threefold: 1) to examine the effect of metacognitive judgments at encoding of related facts on adults' subsequent self-derivation, 2) to test for the potential relation in grade 3, 4, and 5 children's self-derivation performance and their metacognition, and 3) to illuminate age-related change in the relation between self-derivation and metacognition in a longitudinal sample of elementary school-age children across grades 3-5. In this General Discussion, I summarize the major findings relevant to these three goals, explore the implications of the findings, and reflect on some limitations of the work and directions for future research.

Self-derivation through memory integration was the focus of the present research because it stands out from other productive processes such as analogy and associative inference. For example, the process of self-derivation is ecologically valid. In self-derivation, participants learn pairs of facts, information that could readily be learned in the world outside the laboratory, and then derive new knowledge. Furthermore, the knowledge gained through self-derivation is educationally-relevant (Esposito & Bauer, 2017), rapidly incorporated into semantic memory (Bauer & Jackson, 2015), and is retained for at least 1 week (Dugan & Bauer, 2020; Varga & Bauer, 2013), rendering it useful after initial learning. The products of other productive processes are often arbitrary and fleeting (Zeithamova et al., 2012). Crucially, unlike other productive processes (Lazareva & Wasserman, 2010; Schlichting et al., 2014), self-derivation does not require training or instruction for participants to be successful. Some of these characteristics may be true of other productive processes but to date, they are only evident in self-derivation.

Metacognition, the monitoring and control of cognition, likely influences self-derivation performance because it involves understanding one's own learning and taking steps toward better task performance. We specifically hypothesized that accurate metacognition would improve self-derivation performance. There are two routes to this improvement. First, if participants accurately monitor their learning, they may come to realize the related nature of the fact pairs and the integrative structure of the task overall. This knowledge could aid participants in devising a general strategy for greater success at self-deriving. Evidence from adults who engage in multiple blocks of trials that present the opportunity to self-derive (i.e., encode—test; encode—test; etc.) shows that performance improves from the first block to subsequent blocks, despite participants never receiving instructions on how to perform successfully (Menkes, Esposito, & Bauer, 2017). The second route is that if participants' monitoring is accurate and focused on the individual stem facts, then on a trial level, they may use monitoring to better remember the facts necessary for successful self-derivation. This is supported by the multilevel model reflecting that most of the variability in participants' self-derivation performance was at the level of the individual trial. Because monitoring and control processes work in concert, it is reasonable to assume that control also influences self-derivation. Control in the context of self-derivation likely involves deliberate attempts to integrate related stem facts and possibly even extended duration selecting information that is necessary for self-derivation. In a separate test of adults' self-derivation, Varga and Bauer (2017a) consulted participants' knowledge of relatedness between pairs of facts. They found a moderately strong relation between knowledge of relatedness and successful self-derivation, $r = .44$. Furthermore, participants who were unsuccessful on a subset of trials but reported being aware of the opportunity to integrate exhibited longer response times specifically for the unsuccessful trials. This suggests participants

searched their memories longer, for a correct response (Varga & Bauer, 2017a). Because this work represents the first foray into metacognition about self-derivation, we focused only on participants' monitoring throughout the process of self-derivation.

Metacognition as a Tool to Understand and Increase Self-derivation: Study 1

Study 1 provided insight into how metacognition can facilitate self-derivation of new factual knowledge in adults. The findings indicate that for self-derivation trials when college age adults had previously made a judgment of learning (JOL) about one or both of the necessary stem facts, they were more likely to self-derive. Similarly, stem facts that were the subject of a JOL were better remembered than stem facts that were not judged. This suggests that making a metacognitive judgment and considering future memory for a fact contributes to better encoding and memory for that information. It is well established that differential encoding impacts subsequent memory (Becker, Kalpouzos, Persson, Laukka, & Brehmer, 2017). Furthermore, the present data contribute to the currently limited corpus of research specifically examining the effect of making metacognitive judgments on memory outcomes. Mirroring what we observed in the present research, Soderstrom, Clark, Halamish, and Bjork (2015) found that adults' memory for word pairs was higher when the pairs were followed immediately by JOLs. This stands in contrast to work suggesting that any memory improvements following judgments of learning are actually due to retrieval practice (Dougherty, Robey, & Buttacio, 2018; Dougherty, Scheck, Nelson, & Narens, 2005). Crucially, Study 1 brings the field a step closer to understanding metacognition for learning outside the laboratory as the items participants engaged with were novel but true facts.

Study 1 also emphasizes that for metacognition to be efficacious in the context of self-derivation, metacognition should be a process that one actively engages. Metacognition does not

spontaneously interact with other cognitive processes. Actively engaging in metacognitive monitoring can guide one toward success. Metcalfe and Finn (2008) found that when participants selected word pairs to restudy, they chose to study the items they had previously given lower JOLs. Moreover, when and how metacognition is used deserves consideration (Kornell & Bjork, 2007). For example, it is possible to allocate study time to easy items despite knowing that more difficult ones have yet to be studied. In the context of self-derivation one may know that facts are related to each other without using that information to guide learning toward self-derivation.

Adults in Study 1, as in most studies of metacognition, were instructed to provide JOLs for specific facts. They did so. Logically, they could have adopted the instructed strategy of judging their learning and applied it to all trials, even those on which they were not instructed to make JOLs. Logically, this would have resulted in increased learning. It is clear from that data that the adults in this study did not spontaneously extend this strategy. That is, stem facts that were judged were remembered better than those that were not. Though from these findings it appears that adults do not engage their metacognitive abilities unless directed to do so, adults are quite sensitive to even the way they are instructed to provide judgments (Finn, 2008). Finn (2008) found that adults' confidence in their performance differed based on whether judgments were framed as "likelihood they would remember" versus "likelihood they would forget." Findings from Study 1, as well as the literature on metacognition more broadly, suggest that engaging in metacognition can lead to better learning outcomes. At the same time, the findings imply that on average, adults do not engage these processes spontaneously.

Examining Self-derivation and Metacognition across Development: Study 2

Study 2 was the first test of the relation between self-derivation and metacognition between children of grade levels 3-5. We capitalized on an ongoing longitudinal study to ask

participants to make judgments about their self-derivation. Specifically, children in 3rd, 4th, and 5th grades (between-subjects) first predicted how likely they were to choose the correct answer to a forced-choice self-derivation question. Then, after each forced-choice self-derivation and stem fact question, children provided ratings of their confidence in their responses. As was anticipated based on prior research, the older children self-derived more than younger children. Specifically, 4th and 5th graders were successful on more self-derivation trials than their 3rd grade counterparts. Despite not observing significant differences in metacognitive performance across the 3 grades, we found some evidence suggesting metacognition is related to self-derivation performance, even in elementary school students.

Based on previous research of self-derivation performance across development, we expected that children would show age-related increases in their generation of new knowledge. We did observe age-related change in self-derivation performance, to a degree. Fifth graders' self-derivation performance was on par with 3rd graders'. We speculate that this was due in part to the differences in stimuli used for each grade. Previous research demonstrating age-related change used the same stimuli for all age groups (Bauer & Larkina, 2017; Esposito & Bauer, 2017). Different stimuli for each grade were necessary for the larger longitudinal investigation of which Study 2 was a part. In addition to age-related change in self-derivation, we also found a significant correlation between stem fact memory and self-derivation performance across all 3 grade levels; this relation supports that stem fact memory is necessary for successful self-derivation. Indeed, children who remembered more stem facts also performed higher at self-derivation questions.

Longitudinal investigations of metacognition in childhood suggest substantial improvements in metacognitive monitoring between 7 to 11 years of age (Roebers et al., 2019;

Roebers & Spiess, 2017). This evidence led us to hypothesize that if metacognition develops in parallel to self-derivation and is relevant for successfully deriving new knowledge, then self-derivation and metacognition would be related. Contrary to these expectations, we did not observe age-related change in average metacognitive judgments or in the accuracy of metacognitive judgments in the present research. With respect to the accuracy of feeling of knowing judgments, children demonstrated the lowest accuracy compared to their confidence judgments. This is unsurprising given evidence that children are most overconfident in making predictions about their performance and in turn, these judgments are often incorrect (Lockl & Schneider, 2002). Confidence judgments for both self-derivation and stem fact questions would logically be more accurate than FOKs for three reasons. First, confidence judgments are given as an evaluation of performance that has already occurred rather than a prediction of future performance. Second, the self-derivation questions were first asked in open-ended format and then forced-choice format, increasing familiarity with those specific questions. Finally, stem facts were the only information explicitly presented to children, suggesting that they may be most confident in their knowledge of what they have directly experienced as opposed to knowledge they constructed themselves.

Overall, our hypotheses were supported, in part. We found that forced choice self-derivation performance was significantly correlated with the accuracy of confidence judgments immediately following the self-derivation questions. That is, those who performed higher on self-derivation were more accurate in judging their confidence in their performance. Similarly, those who rated their confidence as high for self-derivation questions also were highly confident in their answers to stem fact questions. Though we did not observe significant change in these relations across the three grade levels, these data reflect that accurate metacognition is related to

and perhaps important for successful self-derivation. Additionally, the accuracy of self-derivation confidence judgments was significantly related to the accuracy of feeling of knowing judgments. Feeling of knowing accuracy was not related to self-derivation performance; however, this pattern of relations suggests that when children become more accurate at making predictions about their own performance, they can then extend those predictions in service of successful self-derivation.

We did not observe the change in metacognition for self-derivation we hypothesized would unfold across the 3 grade levels. However, children's metacognitive judgments mirrored what is often observed in the literature. In each grade, participants were the least confident in their feeling of knowing predictions compared to their confidence in their self-derivation and stem fact performance. Average metacognitive judgment value increased by nearly a full point (out of 4) when participants provided confidence judgments for their forced choice self-derivation questions. Confidence in stem fact memory was nominally higher than for self-derivation performance. Though stem fact confidence was not significantly higher than confidence in self-derivation performance, this fits with the current understanding of children's metacognition given that stem facts are directly learned information. Interestingly, children were no more confident nor accurate about their stem fact performance compared to their self-derivation performance suggesting that in terms of children's judgments about their cognitions, self-derived knowledge is not treated differently than is information that is directly learned.

The results of Study 2 must be considered within a broader context. A large body of research demonstrates that children tend to be overconfident in their performance for word pairs or learning new words and their definitions. Unlike in many of these studies, children in the present research learned and made judgments about factual knowledge that they were tasked

with integrating to derive something new. This study was also conducted in a typical elementary school classroom. It is possible that even though they can be accurate at making judgments for word pairs in the laboratory, children may have limited experience using metacognition effectively in their classrooms. Thus, this extends our understanding of children's metacognition in an ecologically relevant context and for building knowledge over time. Metacognition for learning as it occurs outside of the laboratory may be more nuanced than we might expect based on the extant literature.

A cornerstone of Study 2 was the cross-sectional design allowing for investigation of age-related change in self-derivation, metacognition, and the association between the two. We used multilevel modeling to examine the variance explained by age and metacognition while also accounting for the nested structure of the design. Multilevel modeling reflected that there was very little variance between grade levels and also between participants within a given grade level. Approximately 90% of the variability observed within our sample was at the level of the individual trial, rendering it difficult to partition. We postulate that the challenging nature of self-derivation, in comparison to tasks that are frequently used in studies of metacognition, contributed to this individual trial variability.

The findings from Study 2 provide the foundation for Study 3. This study did not allow investigation of the individual variability or continuity in self-derivation or metacognition, or test of whether metacognition influences self-derivation over time. Thus, Study 3 served as a longitudinal study to help elucidate these issues.

Variability and Continuity in Self-derivation and Metacognition over 3 Years: Study 3

In Study 3, we examined both the relation between self-derivation and metacognition in childhood and how that relation may change over the period between grades 3 and 5. As in Study

2, we asked participants to provide predictions of self-derivation performance later in a session and confidence judgments about both self-derivation and stem fact performance.

Study 3 was the first longitudinal investigation of self-derivation. We found significant age-related change in self-derivation performance and stem fact memory. In spite of changes in forced-choice self-derivation performance over time, performance was significantly correlated across years. Indeed, self-derivation performance at Year 1 was the only significant predictor of performance at Years 2 and 3. Each year self-derivation and stem fact performance were significantly correlated with each other. Interestingly, stem fact memory changed across the three years but it was less stable than self-derivation performance within participants. This pattern of relations suggests that age-related change in self-derivation and stem fact memory is driven in part by differential engagement in the subprocesses of self-derivation.

Continuity in metacognition was not as robust as in self-derivation performance. Children's average feeling of knowing judgments did not change over time. Neither did confidence in stem fact memory. Only self-derivation confidence judgments changed significantly over time; children grew less confident in their self-derivation performance. This decrease in confidence is not surprising given prior research suggesting that children become less overconfident about their performance over development (Roebers & Spiess, 2017). Together these results suggest that with time children may begin to develop an understanding that self-deriving new knowledge is more challenging than learning information directly.

Accuracy of metacognitive judgments specifically pertaining to self-derivation, feeling of knowing and confidence judgments, did not change over time. However, accuracy of confidence in stem fact memory did change such that children actually became less accurate over time. It is

unclear what might account for decreased accuracy especially in light of stem facts being the only information participants learned directly.

Intraindividual stability of metacognition was lower than the observed stability of self-derivation performance. Confidence judgments about self-derivation were related year to year, but changed in their relation to stem fact confidence each year. More specifically, the pattern of correlations shifted such that at Year 1 confidence judgments about self-derivation and stem fact performance were related. In Year 2, confidence in self-derivation was related to feeling of knowing judgments but not stem fact confidence. In the final year of the study, confidence in self-derivation was related to both stem fact confidence and feeling of knowing judgments. These changes from year to year suggest that the development of highly accurate metacognition is more protracted than the development of successful self-derivation and that during the period between 8 and 11-years-old, children's metacognition is possibly undergoing changes that impact its stability and its relation to self-derivation. This is further evidenced by a shifting pattern of the relation between metacognitive accuracy and self-derivation performance; accuracy of self-derivation confidence judgments are related to forced-choice self-derivation performance at Years 1 and 3 but not at Year 2.

Overall, the metacognitive results were less robust than the self-derivation findings, but we did find evidence of replication of Study 2 of the relation between self-derivation performance and accuracy of confidence in self-derivation performance. As we observed in Study 2, children who were high performers at self-derivation tended to be accurate in their judgments of their performance, as demonstrated by the significant association between forced-choice self-derivation and the accuracy of the confidence judgments that followed. Furthermore,

in neither study did we find consistent patterns of relations between metacognitive accuracy and self-derivation for a particular age group or timepoint.

Putting Together Knowledge about Self-derivation and Metacognition across Development

Studies of metacognition in development juxtaposed with an investigation of adults' metacognition highlights two factors to consider. First, adults have had more experience being active agents of learning, particularly when learning is oriented to a specific goal. Second, as was demonstrated by the JOL manipulation in Study 1, even adults require some direction to use their metacognitive skills. Children in Studies 2 and 3 were instructed on what metacognitive judgments mean and how to provide them, but it is unclear whether children thoroughly connected their monitoring to the learning in which they were engaged, as would be evidenced by significant and consistent relations between metacognitive accuracy and self-derivation performance. As is suggested by the work of Schneider and Lockl (2002), it is not beyond the realm of possibility that children are accurate in their metacognitions but do not put them into practice.

This series of experiments highlights key points about self-derivation and metacognition across development. Both are reasonably more sophisticated in adults, as we would expect from other research and from the extended experience that adults have compared to children. Importantly, we never provided children or adults with feedback on their performance. Learning how to perform correctly is internally generated, much like the knowledge resulting from self-derivation. Together, Studies 2 and 3 provide compelling evidence that the support offered by metacognition may not be as easy to use for children as it is for adults. Children did not use their metacognitive judgments in a way that benefitted their overall performance. When taken together, the results of these studies readily provide a foundation for future research.

Limitations and Future Directions

The present research was not without limitations. Firstly, a more complete examination of the effect of making judgments at encoding on subsequent self-derivation demands a condition wherein participants provide arbitrary judgments that are not about one's cognition. An arbitrary judgment condition would account for the opportunity to elaboratively encode stem facts present when making a metacognitive judgment. Secondly, the cross-sectional and longitudinal studies were not originally designed to include measures of metacognition. They are also overlapping samples: the participants in Study 3 were a subset of the 3rd grade students in Study 2. These limitations suggest further research is needed to more fully understand how metacognition and self-derivation are related across development. Also with respect to Study 2 and Study 3, we were not able to keep the stimuli the same for participants each year. Each year children learned information that was aligned with their grade-level curriculum but not necessarily equal in difficulty to other grades. Finally, developmental science, particularly longitudinal studies, are prone to missing data. We attempted to account for missing data by using multilevel modeling in both Studies 2 and 3, but because the models were not optimally informative we had to use inferential statistics that are more vulnerable to missing data and unequal numbers of trials.

Future investigations will extend the present research in meaningful ways. We must systematically investigate the impact of metacognitive judgments throughout the process of self-derivation in adults and children. To better understand variability and continuity of metacognition in the context of self-derivation, we must also extend the work of Studies 2 and 3 into older age groups. To fully bridge the gaps in our current understanding, an investigation of metacognition for self-derivation into young adulthood would be particularly illuminating.

Conclusions

Self-derivation is an important process for the accumulation of knowledge. These studies represent the first steps toward understanding metacognition about the process of deriving new knowledge with one's own mind and how self-derivation and metacognition are related across development. Prior to this research, the literature on metacognition for productive processes was quite limited and had not focused on self-derivation of new factual knowledge. The present research accomplished significant milestones in the process of elucidating how metacognition may support the productive extension of knowledge through self-derivation. Our findings suggest that in children metacognition is still developing such that we have limited evidence that self-derivation is supported by metacognition. In contrast, when adults actively engage in metacognitive monitoring during learning, it is clear that self-derivation performance is facilitated. We do not yet know *how* participants might use their metacognition to support themselves successful self-derivation but future research will bring us closer. The present findings call for a systematic investigation of metacognition at each of the steps of ERISS across development to draw a clearer picture of how metacognition relates to engagement in those steps and to pinpoint just where in the process, metacognition may be most effectual for successful self-derivation.