Error Patterns in the Knowledge-Based Inference-Making of Less Skilled Middle-Grade Readers: An Exploratory Study

Amy E. Barth,[1] Cathy Newman Thomas,[2] Heather Kincaid[1], Ethan Ankrum[1], Brandon Ruiz,[1] and Leslie Salazar[1]
[1]Buena Vista University
[2]Texas State University

Abstract

The primary aims of this mixed-method study were to (a) examine the effectiveness of a brief inference intervention, (b) compare the types of knowledge-based inferencing errors less skilled middle grade readers make, and (c) evaluate if self-reported cognitive load relates to inferencing. Participants (*N* = 17) were randomly assigned to a graphic organizer-inference intervention (GO-Inference) (*n* = 9) or business-as-usual (BAU) condition (*n* = 8), and differences between groups were explored for each study purpose. Quantitative and qualitative results suggest that while less skilled readers in the GO-Inference condition made modest progress in forming knowledge-based inferences, they continued to struggle to distinguish relevant from irrelevant information from text and/or retrieve the knowledge necessary to form inferences. Students in the BAU condition were more likely to make errors such as providing irrelevant information or failing to respond. Additionally, students in the GO-Inference condition reported lower cognitive load during inference-making tasks.

Keywords: Knowledge-based inference, reading comprehension, adolescent less skilled readers, cognitive load

More than 80% of eighth-grade students in the United States do not have the content-area knowledge and reading skills they need to enter high school and succeed (ACT, 2019). Further, randomized control trials examining the effectiveness of multi-component reading interventions designed to close this gap among middle-grade struggling readers report effects ranging from small to negligible on standardized measures of comprehension (Scammaca et al., 2016; Scammaca et al., 2015), with more significant effects requiring more than one year of intervention. These findings suggest that (a) current intervention approaches have not been intensive enough to budge the intractable deficits presented by middle-grade struggling readers (see Scammaca et al., 2016; Scammaca et al., 2015) or (b) the components do not adequately build the types of higher-order reading skills (i.e., inference-making, comprehension monitoring, and understanding of text structure) involved in comprehension itself. For this reason, systematic investigations examining the nature of higher order reading skills, such as inference-making, remain a priority in the field. Further, considering students' perceived cognitive load for these higher-order skills can provide important information about the potential effectiveness of interventions.

Knowledge-Based Inferencing

Among adolescent readers, inferencing is the strongest predictor of comprehension after controlling for other reading-related skills (Ahmed et al., 2016). One type of inferences, knowledge-based inferences, is particularly important for comprehension because it helps readers build a meaningful

Barth, A., Newman Thomas, C., Kincaid, H., Ankrum, E., Ruiz, B., & Salazar, L. (2022). Error Patterns in the Knowledge-Based Inference-Making of Less Skilled Middle-Grade Readers: An Exploratory Study. *International Journal for Research in Learning Disabilities, 5*(2), 18-35. https://doi.org/10.28987/ijrld.5.2.18

representation of the text (Kintsch, 1988; O'Brien et al., 2015). Knowledge-based inference-making involves integrating information across sentences and between information in the text with one's existing general knowledge of the topic (O'Brien et al., 2015).

Proficiency in knowledge-based inferencing differs significantly between skilled and less skilled adolescent readers (Barth et al., 2015). That is, less skilled adolescent readers have trouble forming knowledge-based inferences that support understanding of short sections of text (Barnes et al., 2015). Indeed, when required to construct inferences across larger text sections, their accuracy decreases to chance levels (Barth et al., 2015; Cain et al., 2004). Finally, less skilled adolescent readers are less likely, less accurate, and less efficient at forming inferences, particularly inferences that require integrating relevant general knowledge with important information in the text (Barnes et al., 2015; Barth et al., 2015).

Why Might Less Skilled Readers Fail to Form Knowledge-Based Inferences?

First, readers largely fail to form knowledge-based inferences because they lack relevant background knowledge, their knowledge is inaccurate, or their knowledge is incomplete (e.g., Ackerman et al., 1990; Barnes et al., 2015; Cain et al., 2001 ; Casteel, 1993). That is, a knowledge-based inference can only be made when the requisite knowledge to make that inference is available. Second, readers may fail to form knowledge-based inferences because they do not activate, access, or use relevant background knowledge that is stored and available in long-term memory (Barnes et al., 2015; Barnes et al., 1996; Cain et al., 2001). Third, readers may fail to retrieve relevant information from text (Barth et al., 2015) or integrate relevant knowledge with important information in the text (Barnes et al., 2015; Barth et al., 2015; Cain et al., 2001). Fourth, readers may not recognize that an inference is called for to fully understand the text (Whitehead, 1929). Finally, readers may fail to form an inference or may generate an incorrect inference because the cognitive load required to access, retrieve, and integrate relevant background knowledge from semantic memory with important information in the text exceeds their processing capabilities (de Jong, 2010).

Research Informing the Proposed Study

Little research has examined the common sources of inference failure among less skilled readers. To address this gap in the literature, Cain et al. (2001) explicitly trained students in background knowledge to criterion and then examined the effects of background knowledge on inference-making among skilled and less skilled comprehenders age 7 to 8 years old (N = 26). Findings revealed that the errors made by skilled and less skilled comprehenders occurred at different stages in the inference-making process. That is, less skilled comprehenders often failed to accurately recall information that had to be integrated to form the inference. Skilled comprehenders, on the other hand, often failed to accurately integrate the relevant textual premise and knowledge-based item. Finally, a small proportion of errors could be attributable to forming the wrong inference, suggesting that skilled and less skilled comprehenders did a poor job of selecting the relevant information from text and/or from their general knowledge.

Building on the Cain et al. (2001) study, Elbro and Buch-Iverson (2013) examined if use of a graphic organizer designed to help elucidate for readers the textual premise and knowledge-based information that must be integrated to generate an inference led to more accurate inferencing among typically developing sixth-grade students. Following two weeks of daily instruction, students in the graphic organizer condition made significant improvements in inference-making and reading comprehension.

Of importance to the current study is the use and effectiveness of the graphic organizer. The purpose of the graphic organizer was to help students understand that an inference was needed for comprehension and identify the knowledge needed to make the correct inference. The twofold purpose of the graphic organizer is important for several reasons. First, less skilled readers are not strategic in their reading (Pressley, 2002), and graphic organizers have been demonstrated to be an evidence-based practice for students with learning disabilities to improve comprehension, including identification of main idea and development of inferences (Dexter & Hughes, 2011). Second, graphic organizers can provide a scaffold that preserves working memory, thereby reducing cognitive load so that higher-level thinking and reasoning about text can be enabled (Dexter & Hughes, 2010). Third, even when less skilled readers have the knowledge needed to make inferences while reading, they do not generate them as readily as more skilled readers, and in such instances the graphic organizer serves as a prompt. For example, Cain and colleagues (2001) reported that in some cases, students directly answered a question about the knowledge needed to form the inference but did not go on to generate the inference using that knowledge when manded.

A more recent study by Daugaard et al. (2017) reported significant correlations between vocabulary and both comprehension and inference-making among sixth-grade students (*N* = 53). This provides additional support for the notion that knowledge (i.e., word and world) is essential for inference-making among readers in the secondary grades (Ahmed et al., 2016; Cromley & Azevedo, 2007).

Finally, previous research has established the contributions of working memory to inference-making (e.g., Currie & Cain, 2015; Swanson et al., 2018). Inferencing is a complex process that draws on working memory because readers must search, find, retrieve, and then integrate relevant background knowledge from memory with new information extracted from text. Additionally, researchers and theorists are exploring the role of cognitive load and its relation to working memory and reading (de Jong, 2010; Paas & Merriënboer, 2020). Limited working memory suggests a higher cognitive load (Galy et al., 2012), whereas greater working memory capacity suggests a lower cognitive load. For less skilled older readers, considering their perceptions related to the cognitive load for tasks is important because it relates to their willingness to expend effort and engage in complex learning (Stevenson & Mussalow, 2018). Student perceptions of cognitive load can inform instructional design and screen to identify students who are struggling (Laurie-Rose et al., 2014).

The current study builds on this research base in several ways. First, the study used instructional materials frequently used by social studies classroom teachers to teach content but not regularly applied to teach inferencing. Texts were expository to build both content knowledge and word knowledge. This contrasts to prior work by Cain et al. (2001), who used an experimental procedure that taught students a knowledge base about a fictional world prior to measuring their inferencing skills about that planet. Next, like Elbro and Buch-Iverson (2013), a graphic organizer was incorporated into the knowledge-based inferencing instructional condition to help conditionalize the knowledge needed for inference-making among typically developing middle-grade students. Our study extends this work by focusing on less skilled middle-grade readers and identifying their common sources of inference failure.

Finally, we examined cognitive load across the intervention to determine if the intervention reduced readers' perceived burden of forming knowledge-based inferences. No previous inferencing study has addressed perceived cognitive load. Managing cognitive load is critical for freeing up cognitive space for learning, along with the social validity aspect of providing interventions that are perceived by the learners to be useful and acceptable (Wolf, 1978).

Study Purpose

The present study was designed to:

1. Determine the effectiveness of a brief inference intervention for rural, middle-grade struggling readers.
2. Identify whether the inference errors made by students who complete the inference intervention differ from those of students receiving business-as-usual (BAU) instruction.
3. Determine the extent to which the cognitive load of students who complete the inference intervention is lessened relative to students who complete BAU instruction.

To address the first purpose, intervention effectiveness was explored using quantitative methods, including descriptive statistics, to examine demographic characteristics of students, calculation of effect sizes to determine group differences on dependent variables, and use of Mann Whitney U, a non-parametric test to evaluate group differences in inferencing-error types. Given our small sample, case studies were also developed to demonstrate the relationships between standardized test scores, performance on inference-making, and cognitive load to illustrate variables that might be manipulated and guide future research.

To address the second study purpose, student responses were qualitatively coded for errors using a directed content analysis approach (Hsieh & Shannon, 2005) to look for common features based upon and extending the previous early/late staging of errors proposed by Cain et al. (2001). Common error types that emerged during analysis were named and defined to provide a way of interpreting students' skills and quality of inferencing. These were then ordered into a hierarchy of difficulty based on Cain et al.'s (2001) previous work.

Finally, to address the third study purpose, students were asked to complete an adapted form of the NASA Task Load Index (TLX; Hart, 2006), a measure of cognitive load, to learn about their perceptions of the ease or difficulties they experienced when trying to form inferences post intervention. The original quantitative index was developed by NASA to understand how to manage cognitive attention and maximize productivity. For our purposes, it was adapted to be used by children.

Method

Participants

Study participants were students from one middle school (N = 17) serving students in Grades 6-8, located in a rural, working-class community in the midwestern United States. The student population is <5% Asian, 10.8% African American, <5% Hispanic, 0% Indian, and 79.8% White. One hundred percent of students are economically disadvantaged, as measured by participation in the school's free and reduced-price lunch program (see Table 1).

Table 1
Demographics

Variables	BAU	GO-Inference	School District
	(n = 8)	*(n = 9)*	*(n = 345)*
Grade			
6	5	3	130
7	3	3	85
8	0	3	130
Gender			
Male	1	5	177
Female	7	4	168
Free/reduced lunch	8	9	345
Race			
White	8	8	244
African American	0	1	54
Hispanic	0	0	24
Other	0	0	23
Special education			
No	7	8	298
Yes	1	1	47

Note. BAU: Business-as-Usual condition. GO-Inference: Intervention condition.

Potential participants were identified through a two-step process. First, based on scores on the state-mandated reading comprehension competency test, the Missouri Assessment Program (MAP; Missouri Department of Elementary and Secondary Education, 2013), from the previous year; students who performed at or below Basic were identified as less skilled readers and recruited for participation. Second, recruited participants were required to return signed parental consent forms approved by the university's Institutional Review Board and their school district board. Students were excluded from participation if the school identified them as Limited English Proficient or as having a significant disability such as blindness, severe cognitive impairment, or severe behavioral/emotional disability.

All students who consented to participate in the study (N = 17) were randomly assigned to either the GO-Inference condition (n = 9) or the BAU condition (n = 8). The racial/ethnic composition of the final sample was 94% White and 6% African American. Sixty-five percent were female, 100% were eligible for free or reduced-price lunch, and no students received special education services. The final sample included a mix of less skilled readers from Grade 6 through Grade 8, with eight students in Grades 6, six in Grade 7, and three in Grade 8.

Instructional Characteristics and Training

Tutors

Instruction in both the GO-Inference and BAU conditions was provided by two special education doctoral research assistants. These tutors received six hours of training on key instructional elements and procedures, features of effective instruction and behavior management, and strategies for supporting student engagement in both instructional conditions. Additionally, they participated in weekly meetings to receive ongoing instructional support and daily check-ins to ensure high levels of fidelity of implementation.

Duration and Intensity

Students received 20 minutes of instruction 4 times per week for 8 sessions over a 2-week period, for a total of 160 minutes. Instruction was delivered in small, mixed-grade groups. Instructional time, study duration, and release from scaffolding mirror Elbro and Buch-Iverson's (2013) instructional delivery.

Graphic Organizer (GO-Inference) Condition

Tutors utilized a researcher-developed, scripted, explicit, instructional routine and graphic organizer (see Figure 1) to help students acquire knowledge and practice knowledge-based inferencing. Graphic organizers are an evidence-based, high-leverage practice for supporting the development of inference-making (Dexter & Hughes, 2013; Elbro & Buch-Iverson, 2013; McLesky et al., 2017). The intervention consisted of three components: (a) scaffolded instruction, (b) feedback, and (c) gradual release of support.

Component 1 – Scaffolded Instruction. Tutors scaffolded instruction to help students gain knowledge and independence in accurately forming knowledge-based inferences through "think aloud-style" instruction (Kucan & Beck, 1997). The tutor thought out-loud about what they were (a) looking at in text, (b) thinking about relevant knowledge, and (c) doing to integrate information from the text with knowledge from memory. The tutor and students first read the text on ancient Egypt (see Figure 1). Then, the tutor modeled and guided students' formation of knowledge-based inferences using a graphic organizer. The graphic organizer helped to elucidate the information from the text and the general knowledge that should be recalled and then integrated to form knowledge-based inferences about ancient Egypt (see Figure 1).

Component 2 – Feedback. To increase accuracy when forming knowledge-based inferences, tutors gave students written feedback on their graphic organizers. Specifically, the purpose of the feedback was to improve students' accuracy in identifying accurate and relevant information from text or knowledge and then integrating these two pieces of information when forming knowledge-based inferences. For example, the tutor might provide feedback on the accuracy of the prior knowledge, but guide the student to re-examine the paragraph with relevant textual information as students often retrieved proximal information from text rather than the exact information called for. Alternatively, if a student provided no response to prior knowledge, the tutor might supply that knowledge. If the student provided no response to textual information, the tutor might provide cues for where/what to look for. Students used the written feedback to correct any part of their inference (i.e., identification of information from text or knowledge and the integration of text with knowledge).

Component 3 – Gradual Release of Support. The tutor gradually released the amount of instructional support. By the third intervention session, students independently read the passage and inference question, identified important information from text and knowledge required to form the inference, and generated their knowledge-based inference using the graphic organizer. In their small group, with the tutor, students discussed the text and knowledge components of their inference, shared their knowledge-based inference, and used peer feedback and/or written feedback from the tutor to revise any part of their inference.

Business-as-Usual (BAU) Condition

In this condition, tutors used a scripted instructional routine to help students identify the main ideas of informational texts on ancient Egypt. This instructional routine was widely used in the district's English language arts and social studies classrooms to support reading comprehension, was familiar to students, and, therefore, could be considered business-as-usual. In this condition, students were prompted to identify important details in the text and then synthesize them into a main idea statement. Research has supported the contribution of main ideas when forming inferences (Fritschmann et al., 2007), and main idea is a commonly implemented and effective intervention to improve comprehension and build general knowledge (Kim et al., 2012). Instruction consisted of two components: (a) fluent reading and understanding of text and (b) formation of main idea statements.

Component 1 – Fluent Reading and Understanding of Text. To support students' fluent reading and understanding of the text, tutors used an explicit instructional sequence (Archer & Hughes, 2011) of "I do," "we do," and "you do" for each instructional session. Tutors orally read the text, focusing on modeling appropriate fluency, prosody, and pronunciation of vocabulary. Next, they directed the students to whisper-read the passage aloud synchronously. Finally, students read the passage silently.

Component 2 – Main Idea of Text. Following the three readings of the text, the tutor guided students' identification of the main ideas of the text. Small-group discussion helped students to prioritize the most important ideas in the text.

Instructional Materials

For both the GO-Inference and BAU conditions, each instructional lesson consisted of the same 5-9 sentence informational paragraphs ranging in Lexile from 1,000L-1,220L, from the book *Egypt World* (Caldwell, 2013). Each passage included a researcher-developed knowledge-based inference question, which students in the GO-Inference instructional condition answered by completing a graphic organizer (see Figure 1). BAU students completed a worksheet identifying the main idea that would be required if asked to make an inference.

Fidelity of Implementation

Tutors audio-recorded all intervention sessions and completed an implementation fidelity checklist that was specific to their condition. Using the audio-recordings, fidelity checklists were also completed by the primary investigator for four of eight sessions (50%). Checklists documented completion of the intervention components in the GO-Inference and BAU conditions. Using percent agreement, implementation reliability was calculated as 97% for the

GO-Inference condition and 95% for the BAU condition, demonstrating that both interventions were implemented with high fidelity.

Procedures

Students completed a pretest battery designed to describe basic reading skills, verbal and content knowledge, word reading fluency, inference-making, nonverbal reasoning, and working memory. Pretest assessments were completed within a two-week window prior to the first day of intervention. Following the instructional period, students completed a posttest assessment battery designed to examine the types of inference errors middle-grade less skilled readers make. Reading comprehension, verbal knowledge, and content knowledge were also reassessed. Posttest assessments were completed in a two-week window following the last day of intervention. Three months later, a delayed posttest battery was administered to measure the retention of content knowledge and inference-making error types. The delayed posttest assessments were completed in a one-week window. Testing at all three timepoints occurred in the school library during a time identified by the school principal.

Following completion of an extensive training program on test administration and scoring, two graduate research assistants administered all assessments to individual students. After all participants had completed pretest or posttest, research assistants evaluated the fidelity of test administration using a two-step process. First, the research assistants double-checked their own item-level scoring and calculation of raw scores and standardized scores. Second, the research assistants verified each other's item-level scores, raw scores, and standardized scores.

Measures

Reading Comprehension. The Gates-MacGinitie Reading Tests-Fourth Edition (GMRT-4; MacGinitie, 2000) Reading Comprehension subtest is a timed (35-minute), group-administered assessment consisting of expository and narrative passages ranging in length from 3 to 15 sentences. Students read each passage silently and answer multiple-choice questions. Internal consistency reliability ranges from .91 to .93, and alternate form reliability is reported as .80 to .87 (MacGinitie, 2000). The Reading Comprehension subtest was administered at pretest and immediate posttest.

Word Reading Fluency. The Test of Word Reading Efficiency-2 (TOWRE-2; Torgesen et al., 2012) is an individually administered, standardized assessment of word reading fluency. For the Sight Word Efficiency subtest, the participant decodes a list of 104 real words as accurately and efficiently as possible within 45 seconds. For the Phonemic Decoding Efficiency subtest, the participant decodes a list of 63 nonwords as accurately and efficiently as possible within 45 seconds. Alternate form and test-retest reliability coefficients exceed .90 for students in the middle grades. The TOWRE-2 was administered at pretest.

Verbal Knowledge. The GMRT-4 Vocabulary Test (MacGinitie, 2000) is a standardized assessment of verbal knowledge. Students read and answer 45 multiple-choice items that assess word, world, and content knowledge. Internal consistency for students in Grades 6-8 is .83-.89. The GMRT Vocabulary Test was administered at pretest and immediate posttest.

Nonverbal Reasoning. The Kauffman Brief Intelligence Test-2 (K-BIT-2; Kaufman & Kaufman, 2004) Matrices Subtest is a standardized assessment designed to measure fluid thinking. It is individually administered and consists of 46 nonverbal items that involve visual stimuli, both meaningful (i.e., people and objects) and abstract (i.e., designs and symbols). All items are multiple-choice, requiring the participant to point to the correct response or to say its letter. Internal consistency coefficients (split-half) for the nonverbal scores for students in Grades 6 through 8 range from .86 to .91. The Matrices subtest was administered at pretest.

Inference-Making. The Test of Language Competence-2 (TLC-2; Bowers et al., 2009) Listening Comprehension-Inference-Making subtest requires that the student form two plausible inferences on the basis of two sentences that describe the beginning and end of a causal chain. The two inferences are selected from four statements. All 12 items and 2 practice items are read to the student. Internal consistency for the TLC-2 ranges from 0.59 to 0.70. The Listening Comprehension-Inference-Making subtest was administered at pretest.

Working Memory. The Woodcock-Johnson-III (Schrank et al., 2001) Numbers Reversed subtest and Memory for Words subtest are measures of short-term memory. For the Numbers Reversed subtest, students are asked to repeat a series of digits backwards. This subtest requires the ability to temporarily store and recode information presented orally by the examiner. For the Memory for Words subtest, students are asked to repeat a series of unrelated words. The subtest measures verbal memory span. Reliability for both subtests exceeds .90 for students in Grades 6-8. Both working-memory subtests were administered at pretest.

Cognitive Load. An adapted version of the NASA Task Load Index (TLX; Hart, 2006) was implemented to measure perceived cognitive load. The TLX is a self-reported assessment of workload that allows workers to report the demands of tasks related to mental, physical, and temporal demand, performance, effort, and frustration (see https://humansystems.arc.nasa.gov/groups/tlx/). The instrument can be completed online and uses a sliding 7-point scale. For the purposes of this study, participants' ratings of their performance, effort, and frustration were of interest.

We adapted the scale to be completed by paper and pencil and to be age appropriate, creating a checklist for children to indicate the levels of demand they experienced each day of instruction. A 7-point fixed point scale was employed. For performance, students rated their daily work as *failure, not good, needs work, okay, pretty good, very good*, and *perfect*, with *failure* scored as 1 point, and *perfect* scored as 7. For effort and frustration, students rated the demands from *very low, pretty low, a little low, okay, a little high, pretty high,* and *very high*, with *very low* scored as 1 point, and *very high* scored as 7 points. Students were scheduled to complete this measure following daily instruction. Previous research (Laurie-Rose et al., 2014) demonstrated the validity of an adapted NASA TLX self-report of cognitive load with respect to academic tasks, even for very young children.

Content Knowledge. The Egyptian Content Knowledge Assessment (Barth & Elleman, 2017) is a 25-item custom assessment measuring basic knowledge of ancient Egypt. Of the 25 items, 11 are multiple-choice and 14 are constructed response; 6 items tap vocabulary, 6 items tap inferencing, and 13 items tap literal comprehension. Internal consistency, calculated using Cronbach's alpha, was .89 and .85 among students in Grade 5 and Grades 6-8, respectively. The Egyptian Content Knowledge Assessment was administered at pretest, immediate posttest, and delayed posttest.

Proximal Measure of Knowledge-Based Inferencing. Students read two texts (i.e., Nile and Building Pyramids) selected from the Qualitative Reading Inventory-5 (QRI-5; Leslie & Caldwell, 2010) and answered knowledge-based inference questions developed for each text. The Nile passage is 294 words in length, 850 Lexiles in difficulty, and includes 5 knowledge-based inference questions. The Building Pyramids passage is 304 words in length, 850 Lexiles in difficulty, and includes 4 knowledge-based inference questions.

For this task, student read the Nile passage and then orally answered the knowledge-based inference questions. After answering all questions, students received the Graphic Organizer-Inference and were prompted to complete it for each inference question. Next, students completed the Building Pyramids passage using the same format. Internal consistency for Nile and Building Pyramids Total score was .83 (Barth & Elleman, 2017). The Proximal Measure of Knowledge-Based Inferencing was administered at pretest, immediate posttest, and delayed posttest.

Curriculum-Based Measure of Knowledge-Based Inference-Making. The Curriculum-Based Measure (CBM) of knowledge-based inference-making consists of three passages drawn from the book *Egypt World* (Caldwell, 2013). Students read each passage aloud and answered a knowledge-based inference question using the Graphic Organizer-Inference. The graphic organizer permitted examination of the information for text, relevant knowledge, and integration of text with knowledge. The CBM Knowledge-Based Inference-Making measure was administered at immediate posttest and delayed posttest.

Analysis Plan

To address the first study purpose, descriptive statistics for the pretest assessments were first calculated. Second, Cohen's *d* and Hedges' *g* were calculated to quantify how much the treatment group differed from

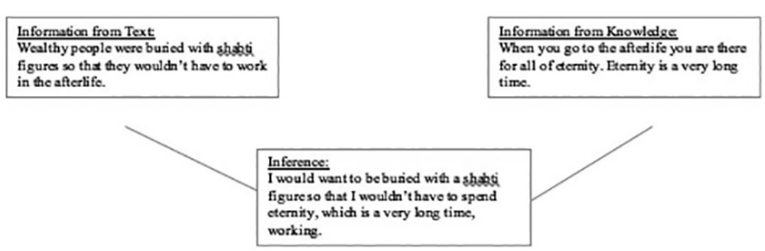

Figure 1

Sample Passage and Graphic Organizer for Knowledge-Based Inferencing

"Ancient Egyptians worried that they might be expected to work in the afterlife. For this reason, wealthy people were often buried with small shabti figures. It was thought that these model servants, inscribed with a spell that allowed them to spring to life when needed, would be able to perform the tasks instead" (Caldwell, 2013). Why would you want to be buried with a shabti figure?

the BAU group on the proximal measures of inference-making following the short intervention period. Both statistics are similar except when sample sizes are below 20 participants, then Hedges' *g* outperforms Cohen's *d* (Durlak, 2009; Hedges & Olkin, 1985).

To address the second study purpose regarding inferencing error types (see Table 2), the Mann Whitney *U*, a nonparametric test, evaluated differences between GO-Inference and BAU conditions on the QRI Knowledge-Based Inference Assessment and CBM Knowledge-Based Inference-Making. The Mann Whitney *U* is a robust test that is appropriate for analysis of ordinal data and accommodates the small sample sizes and non-normal distributions for some data in this study (Nachar, 2008).

A directed content analysis (Hsieh & Shannon, 2005) was conducted to explore the types of errors made by the middle-grade participants. Previous work by Cain et al. (2001) with younger children had staged inference-making errors into (a) early errors by less skilled comprehenders (e.g., failing to recall prior knowledge required) and (b) late errors by more skilled comprehenders (e.g., failure to accurately integrate textual information with prior knowledge). These stages were the starting point for coding in the current study, with the intent of exploring application to older students and also to further define error types. Each error was identified and grouped for similarities by the third author, including all individual student errors across both instructional conditions. Error types were given initial names/codes. The second author independently reviewed each error group for agreement about similarities and naming conventions. The second and third authors met to discuss any differences and complete data reduction and naming. This directed content analysis extended Cain et al.'s (2001) conceptualization of early and late errors to include five types of errors from text, five types of errors from knowledge, and nine types of errors that occurred during the integration of text and knowledge (see Figure 1).

Finally, to address the third study purpose related to the cognitive load associated with knowledge-based inference-making, we compared the means on the three dimensions of cognitive load – performance, effort, and frustration. Also, because no study to date has examined the relation between inferencing and cognitive load, data from individual case studies of student participants were examined to elucidate any trends that may inform future research and future intervention design.

Table 2
Knowledge-Based Inferencing Error Types

Category of Error	Specific Type of Error
Text	Inaccurate Information From Text
	Irrelevant Information From Text
	Text Information Replaced With Knowledge Information
	Omission of Information From Text
	Total Text Errors
Knowledge	Inaccurate Information From General Knowledge
	Irrelevant Information From General Knowledge
	General Knowledge Replaced With Text Information
	Omission of General Knowledge
	Total Knowledge Errors
Integration of Text and Knowledge	Inaccurate Text Information Used During Integration
	Irrelevant Text Information Used During Integration
	Omission of Text Information During Integration
	Inaccurate General Knowledge Used During Integration
	Irrelevant General Knowledge Used During Integration
	Omission of General Knowledge During Integration
	Total – Text Integration Errors
	Total – Knowledge Integration Errors
	Totals – Knowledge and Text Integration Errors

Results

Descriptive Statistics

Table 2 reports demographic information for the GO-Inference and BAU conditions and the participating school district.

Table 3 reports descriptive statistics on standardized measures of word reading fluency, vocabulary, reading comprehension, nonverbal intelligence, inference-making, and working memory for students randomized into GO-Inference and BAU conditions. As seen when examining pretest performance, no significant differences existed between the BAU and GO-Inference intervention groups at pretest. Participants performed approximately one standard deviation below the mean on a standardized measure of word reading fluency and approximately two thirds of a standard deviation below the mean on a standardized measure of reading comprehension, nonverbal reasoning, and working memory. Students performed at the 50th percentile on the standardized measure of inference-making. Also, students performed close to the mean on a standardized measure of verbal knowledge but generally knew less than half of the Egyptian content knowledge before instruction began (see Table 3).

Table 3
Pretest Descriptive Statistics

Assessment	GO-Inference		BAU	
	Mean	*SD*	*Mean*	*SD*
TOWRE	84.78	12.03	80.25	4.68
GMRT Vocabulary	99.94	13.43	100.39	12.77
GMRT Comprehension	92.72	10.38	90.15	11.83
KBIT-2 Matrices	88.67	19.14	87.13	14.02
TLC-2 Listening Comp – Inference-Making	9.11	1.83	10.63	1.51
WJ Memory for Words_SS	89.0	11.19	88.88	22.56
WJ Memory for Words_W	484.11	12.30	480.50	32.85
WJ Numbers Reversed_SS	83.89	10.47	87.34	27.78
WJ Numbers Reversed_W	491.0	12.0	491.20	31.01

Note. *Group differences $p<.05$. SS = standardized score; W = w-score.

Study Purpose 1: To determine the effectiveness of a brief inference intervention for rural, middle-grade struggling readers.

Table 4 reports performance on the measures administered across the three testing time points (i.e., pretest, immediate posttest, and delayed posttest). Hedges' *g* suggests that at the immediate posttest, very small differences existed between the BAU and GO-Inference groups on the Egyptian Content Knowledge total score ($g = 0.04$). Looking closely at the literal items on the assessment, we can see that the BAU condition, which focused on generating main idea statements, outperformed the GO-Inference condition ($g = -0.39$). In contrast, on the inference items, the GO-Inference condition performed slightly higher than the BAU ($g = 0.14$). At the delayed posttest, the literal and inference scores favored the GO-Inference condition ($g = .18$ and $g = .14$, respectively).

On the Proximal Measure of Knowledge-Based Inference-Making, the GO-Inference condition outperformed the BAU condition on the total raw score ($g = .86$), text score ($g = .80$), knowledge score ($g = .39$), and integration score ($g = .62$). Students in both conditions performed near the floor on knowledge and integration at pretest. A similar pattern was found at the delayed posttest, with effect sizes ranging from $g = 0.09$ to 0.81.

The CBM Knowledge-Based Inference-Making, comprised of texts from the intervention, was administered at the posttest and delayed posttest. At posttest, students in the GO-Inference condition performed higher than those in the BAU condition on the total raw score ($g = 0.27$), text score ($g = 0.51$), and knowledge score ($g = 0.24$), but lower than BAU on the integration score ($g = -.60$). At the delayed posttest, the GO-Inference condition again outperformed the BAU on the raw score ($g = .55$), text ($g = .51$), and knowledge score ($g = 0.16$) but now also on the integration score ($g = 0.65$).

Study Purpose 2: To identify whether the inference errors made by students who complete the inference intervention differ from those of students receiving BAU instruction.

To examine whether the types of inference errors differed by condition, student responses on the Proximal Measure of Knowledge-Based Inference-Making and CBM Knowledge-Based Inferencing were combined and analyzed with respect to the 19 potential errors that could lead to an incorrect knowledge-based inference. Errors were classified as coming from text, knowledge, or an integration of the two (see Table 2). We determined if the text-based and knowledge-based information was inaccurate, irrelevant, omitted, or if students provided information from text when they should have provided general knowledge (and vice versa). Table 5 reports significant differences between GO-Inference and BAU conditions.

When the error scores from the two measures, Proximal Measure of Knowledge-Based Inference-Making and CBM Knowledge-Based Inferencing, were combined, four errors were found to be significantly different between the two groups: (a) Irrelevant Information From Knowledge, (b) Irrelevant Information From Text Used During Integration, (c) Omission of Information From Text Used During Integration, and (d) Irrelevant Information From Knowledge Used During Integration.

First, students in the GO-Inference condition tended to provide significantly more irrelevant information from knowledge when forming knowledge-based inferences than students in the BAU

Table 4
Summary of Descriptive Statistics for Pretest, Posttest, and Delayed Posttest

Measure	Condi-tion	Pretest		Posttest		Comparison		Delayed Posttest		Comparison	
		M	*SD*	*M*	*SD*	*d*	Hedges' *g*	*M*	*SD*	*d*	Hedges' *g*
Gates Vocab	GO	99.94	13.43	93.72	11.02						
	BAU	100.39	12.77	99.14	6.96						
						-0.617	-0.55				
Gates Comprehension	GO	92.72	10.38	92.5	10.07						
	BAU	90.15	11.83	92.01	12.88						
						0.045	0.041				
Content Knowledge RS	GO	10	2.35	11.22	3.15			10.89	3.18		
	BAU	11.13	3.98	11.63	2.83			9.25	4.65		
						-0.145	-0.13			0.444	0.396
Content Knowledge VS	GO	2.22	0.83	1.89	0.78			2.22	1.09		
	BAU	1.38	0.92	1.5	0.76			1.38	0.74		
						0.539	0.48			0.948	0.846
Content Knowledge LS	GO	5.33	3.25	6.11	2.56			5.67	2.55		
	BAU	7.13	2.33	7	1.6			5.13	3.23		
						-0.436	-0.389			0.199	0.178
Content Knowledge IS	GO	2.44	1.13	3.22	1.39			3	1		
	BAU	2.63	1.41	3	1.6			2.75	2.25		
						0.157	0.14			0.156	0.139
QRI Raw Score	GO	3.78	2.22	9.67	3.91			9.56	5.2		
	BAU	3	2.39	6.13	3.87			6.13	4.26		
						0.968	0.863			0.763	0.68
QRI Text Score	GO	3.11	1.7	5.67	2.06			5.22	3.5		
	BAU	2.16	5.67	3.63	2.77			2.39	3.07		
						0.898	0.801			0.911	0.812
QRI Knowledge Score	GO	3.11	0.71	2.22	1.48			2.33	2.13		
	BAU	0.88	1.46	1.63	1.41			2.13	2.36		
						0.434	0.387			0.095	0.085
QRI Integration Score	GO	0.33	0.71	1.78	1.4			2	0.5		
	BAU	0	0	0.85	1.46			1.73	0.76		
						0.693	0.618			0.453	0.404
CBM Raw Score	GO	n/a	n/a	2.44	1.74			2	1.5		
	BAU	n/a	n/a	2	1.31			1.25	1.04		
						0.301	0.269			0.612	0.545
CBM Text Score	GO	n/a	n/a	1.56	1.24			1.44	0.88		
	BAU	n/a	n/a	1	0.76			1	0.76		
						0.571	0.509			0.567	0.505
CBM Knowledge Score	GO	n/a	n/a	0.78	0.67			0.33	0.5		
	BAU	n/a	n/a	0.63	0.52			0.25	0.46		
						0.264	0.235			0.177	0.158
CBM Integration Score	GO	n/a	n/a	0.11	0.33			0.22	0.44		
	BAU	n/a	n/a	0.38	0.52			0	0		
						-0.67	-0.597			0.729	0.65

Note. Content Knowledge RS = Egyptian Content Knowledge Assessment Raw Score. Content Knowledge VS = Egyptian Content Knowledge Assessment Vocabulary Score. Content Knowledge LS = Egyptian Content Knowledge Assessment Literal Comprehension Score. Content Knowledge IS = Egyptian Content Knowledge Assessment Inferencing Score.

condition ($U = 14$, $p = 0.018$). Second, compared to the BAU condition, the GO-Inference condition integrated irrelevant information from text when forming knowledge-based inferences ($U = 17$, $p = 0.037$). Third, relative to the GO-Inference condition, the BAU condition failed to provide any information from text when forming knowledge-based inferences ($U = 6.5$, $p = 0.001$). Fourth, when asked to integrate information from knowledge, the students in the GO-Inference condition provided more irrelevant information than the students in the BAU condition ($U = 13.5$, $p = 0.0135$).

Table 5

Significant Differences in Errors Made While Forming Knowledge-Based Inferences by GO-Inference and BAU Condition

Error Type	MWU	*p*	MDN	Range	M.R. Graphic Organizer-Inference	M.R. BAU
Error 1	14	0.02	-1	-2 - 2	11.44	6.25
Error 2	17	0.04	0	-3 - 3	11.11	6.63
Error 3	6.5	0.001	2	-2 - 11	5.72	12.69
Error 4	13.5	0.01	0	-2 - 1	11.50	6.19

Note. $p < .05$ MWU = Mann Whitney *U*. MDN = median. M. R. = Mean Rank. Error 1: Irrelevant General Knowledge Used During Integration on Combined Measure (i.e., Proximal Measure of Knowledge-Based Inferencing Plus CBM Knowledge-Based Inference-Making). Error 2: Irrelevant Text Information Used During Integration on Combined Measure. Error 3: Omission of Text Information During Integration on Combined Measure. Error 4: Irrelevant General Knowledge Used During Integration on Combined Measure.

Study Purpose 3: To determine the extent to which the cognitive load of students who complete the inference intervention is lessened relative to students who complete BAU instruction.

To examine if the cognitive load associated with knowledge-based inference-making varied by condition, we examined the means of the self-reported measure of cognitive load. Descriptive results showed that students in the experimental GO-Inference condition were more confident in their inference-making skills, rating their performance an average of 6.67 on a 7-point scale, compared to students in the BAU condition, who rated their performance an average of 4.73. Students in the latter condition felt that making inferences was more effortful than their peers in the GO-Inference condition, ranking their effort at 3.28, between *a little low* and *okay*. Participants in the GO-Inference condition rated their effort between *pretty low* and *a little low* at 2.5. Students in the GO-Inference condition rated their frustration at 1.67, between *very low* and *pretty low*, while students in the BAU condition rated their frustration a bit higher at 3.32, between *a little low* and *okay*. (See Figures 2 through 4 for comparisons between BAU and GO-Inference groups on the cognitive load measures of performance, effort, and frustration, respectively. See Table 6 for mean scores by participant.)

Student Case Studies

Given the exploratory nature of this research, data from individual case studies of student participants may contribute to understanding of the patterns in the data. Specifically, we examined the reading, cognitive, and cognitive-load data for two students from each condition. In the Go-Inference condition, we

Table 6

Case Study Participant Scores by Condition for Cognitive Load, KBIT-2, TOWRE, and WJ-IV

Participant	Condition	Measure								
		NASA Task Load Index			KBIT	WJ		TOWRE	Gates-MacGinitie	
		Performance	Effort	Frustration	VK	MW	NR	Composite	Vocab	Comp
Participant 7	GO	7	2.71	1	88	82	84	80	87.39	87.4
Participant 15	GO	6.875	1.125	1	104	83	68	66	77.39	77.39
Participant 4	BAU	4.7	3.86	3.57	98	52	90	84	100.3875	87.4
Participant 14	BAU	5.14	5	5.14	97	117	118	84	115.3875	83.387

Note. *pretest scores. NASA Task Load Index of cognitive load provides raw-score estimates of Performance, Effort, and Frustration. KBIT VK: Kauffman Brief Intelligence-2 Verbal Knowledge standard score. WJ MW: Woodcock Johnson-III Memory for Words standard score. WJ NR: Woodcock Johnson-III Numbers Reversed standard score. TOWRE Composite is the Inference-Makings combined standard score for Sight Word Reading Efficiency and Phonemic Decoding Efficiency subtests. Gates MacGinitie Vocab represents the Vocabulary standard score. Gates-MacGinitie Comp represents the Comprehension standard score.

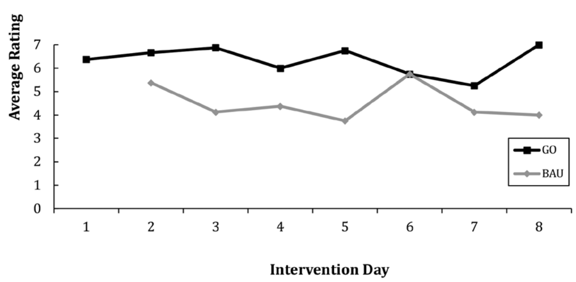

Figure 2
Changes in Average Performance Rating Through Intervention Days

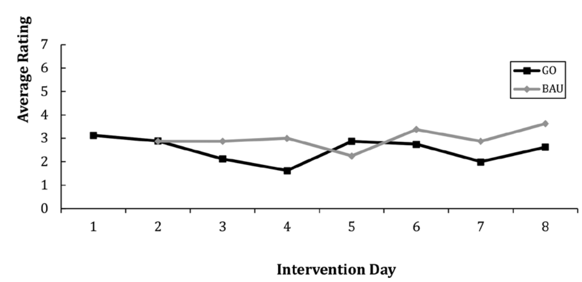

Figure 3
Changes in Average Effort Rating Through Intervention Days

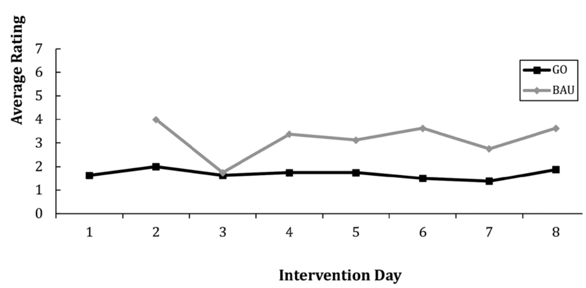

Figure 4
Changes in Average Frustration Rating Through Intervention Days

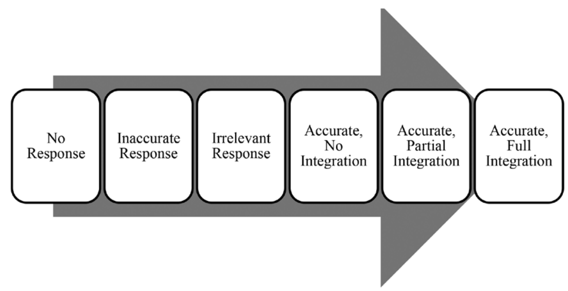

Figure 5
Hypothesized Stages of Errors Made When Forming Knowledge-Based Inferences

examined the performance of Participants 7 and 15. In the BAU condition, we examined the performance of Participants 4 and 14. Below we summarize important information for the selected participants.

Participant 7 (GO)

Participant 7 had a standard score of 88 on the KBIT-2 Verbal Knowledge, representing average potential and indicating average general knowledge. On tests of working memory, standard scores were 82 for Memory for Words and 84 for Numbers Reversed, falling just below average in each case, leading us to anticipate some weaknesses in working memory. The TOWRE standard score of 80 demonstrates sight word and non-sense word reading below average. The Gates-MacGinitie standard scores for vocabulary and comprehension were in the average range. On the cognitive load measure, this student reported confidence in their performance with a score of 7/7, pretty low effort with a score of 2.71/7, and little frustration with a score of 1/7. On Instructional Day 3, in response to a prompt about the Nile being the river of life, this student reported from text that "it had rich black mud," from knowledge that "fertilized land/soil helps to grow crops," making the inference that "the Nile flooded leaving the rich black soil."

Participant 15 (GO)

Participant 15 had a standard score of 104 on the KBIT-2 Verbal Knowledge, falling just above the mean. The participant's standard scores on the tests of working memory were 83 for Memory for Words, just slightly below average, and 68 for Numbers Reversed, indicating a more significant weakness for working memory. The below-average standard score of 66 on the TOWRE indicates that decoding and sight word reading skills are limited. Vocabulary and comprehension as measured by the Gates-MacGinitie were also low, each with scores of 77. In spite of average intelligence as measured by KBIT-2, limitations in working memory and general knowledge are expected to impact reading performance. On the self-report of cognitive load, this student rated their performance at 6.875 on a 7-point scale, almost *perfect*, and rated both effort (1.125/7) and frustration (1/7) as *very low*. On Instructional Day 5, prompted to make inferences about why Thutmose took revenge on his mother, from the text, the student reported "so she would not be king," and from knowledge, "so her name would not be remembered in history," with the inference "so he could get power and rule Egypt."

Participant 4 (BAU)

For Participant 4, the KBIT-2 Verbal Knowledge score of 98 approached the mean. For working memory, there was a discrepancy between the subtests, with a very low standard score of 52 for Memory for Words and an average score of 90 for Numbers Reversed. Numbers Reversed requires the ability to hold and manipulate numbers in a sequence. Familiarity with numbers and order aids in this task. Memory for Words may compound demand because unrelated words must be manipulated and reported in a correct, new sequence. Both tasks rely on auditory attention and working memory. The weakness in Memory for Words might impact the ability to organize, store, and retrieve sight words and map fluently in decoding. The TOWRE-2 standard score of 84 is just below average for sight word and non-sense word reading. Based on Gates-MacGinitie standard scores, vocabulary skills were at the mean, and comprehension skills were average. On the self-report of cognitive load, this student reported their skills as falling between *okay* and *pretty good* (4.7/7), and reported effort (3.86/7) and frustration (3.57/7) between *a little low* and *okay*.

Participant 14 (BAU)

For Participant 14, the KBIT-2 Verbal Knowledge standard score of 97 approximates the mean. This student had the strongest scores across all participants for working memory, with scores just above average of 117 and 118, respectively, for Memory for Words and Numbers Reversed. Comparably, Gates-MacGinitie vocabulary was 115, at the top of the average range. However, comprehension was 83, falling just below average. TOWRE-2 standard scores estimating sight word reading and non-sense word reading just below average may be related to comprehension limitations. Further, this student reported confidence in their performance as *pretty good* (5/7) and rated their effort (5/7) and frustration (5.14/7) both as *a little high*.

Discussion

In this exploratory study, instruction was designed to promote acquisition of relevant knowledge and to provide opportunities to form knowledge-based inferences. Less skilled adolescent readers are less efficient at forming knowledge-based inferences relative to proficient adolescent readers (Barnes et al., 2015) even when they possess the knowledge needed to form the inference. Inference failure may be caused by several reasons, such as a failure to recall and/or integrate information from the text with relevant knowledge of the topic. In addition, readers may generate incorrect inferences rather than those intended or may fail to make an inference because they do not realize that one is necessary to maintain their understanding of text (Cain et al., 2001).

A failure to form knowledge-based inferences for any one of these reasons will result in a less detailed and integrated understanding of text and impair comprehension. As a result, it is important to understand how to improve the accuracy with which inferences are made, the types of errors students make following instruction, as well as students' perceptions of instruction and their performance. Information about perceived cognitive load can be informative for instructional design and delivery to reduce cognitive load. Further, managing cognitive load can increase student engagement and self-efficacy for learning, factors that are especially important for middle-grade struggling readers (Billingsley et al., 2018; Guthrie & Davis, 2013). Below, we summarize and discuss the results of this study as they relate to these points.

Key Findings

Effectiveness of a Brief Inference Intervention

Following eight intervention sessions (i.e., 160 minutes of instruction), students in the GO-Inference condition outperformed students in the BAU condition on multiple measures of knowledge-based inference-making. Very small but practically meaningfully differences were found on a proximal measure of Egyptian content knowledge and a standardized measure of reading comprehension, both favoring the GO-Inference condition. Results align with those of Elbro and Buch-Iverson (2013) showing large, positive effects of a graphic organizer-based inference intervention for typically developing middle-grade students following 240 minutes of instruction (i.e., eight, 30-minute sessions). It is important to note that the primary difference between Elbro and Buch-Iverson's and the present research is the magnitude of the effect sizes.

Our results suggest that a graphic organizer approach that proceduralizes the inferencing process holds potential for improving inference-making among less skilled middle-grade readers. An important next step is to determine the duration and intensity necessary to both automatize the inferencing process and ensure that less skilled readers form as many inferences while reading as more skilled readers do.

Inferencing Errors

Results of this study also suggest that after providing less skilled adolescent readers with the opportunity to learn a knowledge base on ancient Egypt and practice in the formation of knowledge-based inferences, they had trouble discriminating between relevant vs. irrelevant but related background knowledge needed to form knowledge-based inferences. That is, when failing to form knowledge-based inferences, students in the GO-Inference condition correctly identified related information from text but incorrectly identified the knowledge that should be integrated to form the inference. This is encouraging in that they recognized that an inference was called for, and their knowledge was conditionalized and activated, although they were not able to accurately select the exact information from memory that was needed. This is consonant with the general lack of specificity less skilled readers evidence in word knowledge and finding main ideas (Hogan et al., 2011; Scammaca et al., 2007). In comparison, students in the BAU condition often offered no response, either indicating that they had not learned the knowledge or it was not activated when called for.

Interesting, despite daily exposure, neither group grew significantly in their content knowledge of ancient Egypt, making the attempts by the GO-Inference group to provide the requisite knowledge to form inferences even more noteworthy. Evaluation of the knowledge provided by students in the GO group revealed that the knowledge was not completely wrong. It consisted of information that was topically related but irrelevant for an accurate inference to be made.

At the point of integration, students in the Go-Inference condition failed to form correct knowledge-based inferences for two reasons. First, if the knowledge they identified was irrelevant, they consistently integrated this irrelevant knowledge into their inference. Second, in some cases, students identified the correct textual premise only to integrate irrelevant information from text with general knowledge. Collectively, these results suggest that the difficulties less skilled adolescent readers face when forming knowledge-based inferences occur at the point of knowledge identification and continue throughout the integration process (Barnes et al., 2015; Cain et al., 2001). Further, they suggest that learning about ancient Egypt and the inferencing process was occurring across the intervention and that additional modeling of the inferencing process and opportunity to acquire the requisite knowledge and use it to form inferences has potential for improving inferencing and comprehension outcomes.

Results of this study also indicate that relative to the GO-Inference condition, students in the BAU condition failed to integrate relevant information from text needed to form knowledge-based inferences. That is, although students correctly identified important information from text, they omitted this information upon integration with general knowledge, often offering no response at all. These findings suggest that main idea approaches designed to teach readers how to identify important information in text, but not to understand when and why to use that knowledge should be useful. Thus, main idea instruction in this study did not generalize to inference-making among less skilled adolescent readers, who needed to both identify and integrate relevant textual premises with relevant knowledge when forming knowledge-based inferences. This is supported by recent research showing that main idea approaches do not generalize to improved inferencing among less skilled adolescent readers (Barth et al., 2016).

Based on the types of errors less skilled adolescent readers made when forming knowledge-based inferences, our qualitative data suggest that students appear to progress through stages as they learn to accurately form knowledge-based inferences (see Figure 5). This notion aligns with previous work by Cain et al. (2001) indicating that less skilled readers make inferencing errors at an earlier stage in the inferencing process than more skilled readers.

Our data suggest that regardless of condition, both groups of students progressed through the same stages. However, students in the GO-Inference condition transitioned across a greater number of stages than students in the BAU condition. This suggests that explicit instruction that specifically teaches the process of inferencing and provides targeted corrective feedback helps students to operationalize the process of inferencing. The proposed stages are based on the performance of a small sample of less skilled adolescent readers, and future quantitative and qualitative research is required to validate these stages with a larger, more diverse sample. Our findings do, however, provide a starting point for a finer-grained analysis of errors that can inform instruction and guide intervention.

Our work suggests that in Stage 1, students fail to provide information from text or general knowledge. Stage 2 students provide information from text or general knowledge, but one or both pieces are inaccurate. Stage 3 students provide information from text and knowledge, but one or both pieces are irrelevant. Stage 4 students accurately identify the relevant information from text and general knowledge but fail to accurately integrate both pieces of information. Stage 5 students accurately identify but only partially integrate information from text and general knowledge. Finally, Stage 6 students accurately identify and fully integrate information from text and general knowledge.

Cognitive Load

Additionally, our findings suggest that future inference-making research should consider the perceived cognitive load of this complex task for students and its relationship to learning outcomes. In this study, students in the experimental (GO-Inference) condition rated their performance higher and their effort and frustration as less than their peers in the BAU condition, suggesting that the explicit instruction in inference-making and the scaffold of the graphic organizer helped to reduce the demands on working memory and to manage extrinsic cognitive load.

Managing cognitive load is intended to optimize long-term memory (Chandler & Sweller, 1991). For inference-making, which relies on the efficient search and retrieval of specific information from long-term memory (O'Brien et al., 1998), considering cognitive load in instructional design may support improved learning. With this small sample, examining relationships among cognitive skills such as working memory and cognitive load was not possible, but should be studied in the future, along with additional studies to validate adapted measures of cognitive load for children (Laurie-Rose et al., 2014).

Practical Implications

This experimental trial, using randomization to condition, was designed to improve our understanding of the types of errors less skilled adolescent readers make when forming knowledge-based inferences. The practical implication of these findings suggests first that less skilled adolescent readers experience difficulty identifying relevant information from text or general knowledge needed to form knowledge-based inferences. Consequently, instructional approaches designed to improve inferencing need to explicitly teach students how to both identify and integrate relevant information. Graphic organizers demonstrate some evidence for capacity to scaffold this knowledge activation and integration.

Second, the results suggest that main idea approaches that help readers to identify important information in text, but do not explicitly show students how to integrate important information in text when forming knowledge-based inferences, may not be sufficient for supporting the formation of knowledge-based inferences among less skilled adolescent readers.

Third, as students are learning how to form knowledge-based inferences, they may transition in the types of errors that they make. Consequently, classroom teachers should examine the process of inferencing (i.e., types of errors) as much as the product of inferencing (i.e., was the inference correct or incorrect?). An important next step is to determine how much instruction and practice is required to move students through the error process more rapidly as well as the type of feedback that facilitates this transition. This step is clinically important because it will provide insight about the duration and intensity of instruction needed to help adolescent less skilled readers form knowledge-based inferences during reading as well as the type of feedback that teachers might use to move students across stages.

Fourth, for older struggling readers, who have often experienced substantial and persistent academic failures (Stevenson & Mussalow, 2018), and who have cognitive profiles that predict poor reading performance *and* relate to behaviors such as disengagement, avoidance, and other challenging behaviors, considering their perceptions of instructional conditions can (a) guide instructional design (Billingsley et al., 2018; Guthrie & Davis, 2013), (b) be useful in screening for early intervention (Laurie-Rose et al., 2014), and (c) serve as a potential antecedent preventative for behavior (Stevenson & Mussalow, 2018).

Study Limitations and Future Research

This study provides preliminary information about the nature of inferencing errors among adolescent less skilled readers. The exploratory nature of the study revealed a couple of important limitations to consider for future research. First, although two measures of knowledge-based inferencing were administered, future research should further develop and refine measures of knowledge-based inference-making. For example, readers make a variety of knowledge-based inferences while reading. Inferences may be generated to create causal, spatial, or temporal relations or be made to establish intensions, motivations, emotions, or traits important for maintaining coherence (Hall & Barnes, 2017). Given the wide variety of inferences readers make, inference assessments that more broadly assess these different types could serve to isolate the different points of inference breakdown. Such data could then be used to create more comprehensive inference interventions.

Second, instruction in the current study was only 160 minutes long and designed to promote knowledge acquisition and provide practice in forming knowledge-based inferences. The short duration is consonant with other studies of knowledge-based inference-making (Cain et al., 2001; Elbro & Buch-Iverson, 2013). However, reading research has established that less skilled adolescent readers are often resistant to intervention and require increased intensity and duration to influence learning (Denton et al., 2013). Future research and practice, therefore, should increase the duration, frequency, and intensity of inference instruction.

Third, future studies should explore multiple intervention components. In the present study, the graphic organizer did cue knowledge retrieval, but the knowledge test demonstrated insufficient knowledge development at posttest, and vocabulary knowledge specific to the curriculum was not assessed. Including explicit instruction in each of the variables that diminish inference-making capacity should be included and evaluated in future multi-component reading interventions.

Fourth, our sample size was small. Although students were randomized to condition, future research should increase the sample size such that it is fully powered to test the effects of intervention. Given the small sample size, for example, it was not possible to observe developmental differences across grades or differences perhaps attributable to demographic variables such as free and reduced-price lunch status.

Fifth, the sample lacked diversity. Future research should diversify the demographics of the sample to better understand how interventions designed to promote knowledge acquisition and teach knowledge-based inferencing generalize across the various subgroups of less skilled readers in the middle grades.

Sixth, research should further explore the hierarchy of error types proposed by this study, as well as develop competing models for consideration. Finally, an adapted version of the cognitive load measure (TLX) was utilized in this study. Future studies might consider using the original measure, with a larger sample size, that is fully powered to detect differences between the intervention and BAU groups.

References

Ackerman, B. P., Silver, D., & Glickman, I. (1990). Concept availability in the causal inference of children and adults. *Child Development, 61,* 230-246. https://doi.org/10.2307/1131062

ACT. (2019). *The condition of college and career readiness.* www.act.org/condition2019

Ahmed, Y., Francis, D., York, M., Fletcher, J., Barnes, M., & Kulsz, P. (2016). Validation of the direct and inferential mediation model of reading comprehension among students in grades 7 through 12. *Contemporary Educational Psychology,* 44-45, 68-82. https://doi.org/10.1016/j.cedpsych.2016.02.002

Archer, A., & Hughes, C. A. (2011). *Explicit instruction: Effective and efficient teaching*. Guilford Press.

Barnes, M. A., Ahmed, Y., Barth, A., & Francis, D. J. (2015). The relation of knowledge-text integration processes and reading comprehension in 7th- to 12th-grade students. *Scientific Studies of Reading, 19*(4), 253-272. https://doi.org/10.1080/10888438.2015.1022650

Barnes, M. A., Dennis, M., & Haefele-Kalvaitis, J. (1996). The effects of knowledge availability and knowledge accessibility on coherence and elaborative inferencing in children from six to fifteen years of age. *Journal of Experimental Child Psychology, 61*, 216-241. https://doi.org/10.1006/jecp.1996.0015

Barth, A. E., Barnes, M., Francis, D., Vaughn, S., & York, M. (2015). Inferential processing among adequate and struggling adolescent comprehenders and relations to reading comprehension. *Reading and Writing, 28*(5), 587-609. https://doi.org/10.1007/s11145-014-9540-1

Barth, A. E., & Elleman, A. (2017). Evaluating the impact of a multi-strategy inference intervention for middle-grade struggling readers. *Language, Speech, and Hearing Services in Schools*, *48*(1), 31-41. https://doi.org/10.1044/2016_lshss-16-0041

Barth, A. E., Vaughn, S., Capin, P., Cho, E., Stillman-Spisak, S., & Martinez, L. (2016). The effects of a text-processing comprehension intervention on struggling middle school readers. *Topics in Language Disorders, 36,* 368-389. ttps://doi:10.1097/TLD.0000000000000101.

Billingsley, G., Thomas, C. N., & Webber, J. (2018). Effects of student choice of instructional method on the learning outcomes of students with comorbid learning and emotional/behavioral disabilities. *Learning Disability Quarterly, 41*(4). https://doi.org/10.1177/0731948718768512

Bowers, L., Huisingh, R., & LoGiudice, C. (2009). *The Listening Comprehension Test—Adolescence*. LinguiSystems, Inc.

Cain, K., Oakhill, J. V., Barnes, M. A., & Bryant, P. E. (2001). Comprehension skill, inference-making ability, and their relation to knowledge. *Memory & Cognition*, *29*(6), 850-859. https://doi.org/10.3758/BF03196414

Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology*, *96*(1), 31-42. https://doi.org/10.1037/0022-0663.96.1.31

Caldwell, S. (2013). *Egypt world: Discover the wonders of the ancient land of Tutankhamun and Cleopatra.* Scholastic.

Casteel, M. A. (1993). Effects of inference necessity and reading goal on children's inferential generation. *Developmental Psychology*, *29*(2), 346. https://doi.org/10.1037/0012-1649.29.2.346

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, *8*(4), 293-332. doi:10.1207/s1532690xci0804_2

Cromley J. G., & Azevedo, R. (2007). Testing and refining the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology*, *99*, 311-325. doi:10.1037/0022-0663.99.2.311.

Currie, N. K., & Cain, K. (2015). Children's inference generation: The role of vocabulary and working memory. *Journal of Experimental Child Psychology, 137*, 57-75. doi:https://doi.org/10.1016/j.jecp.2015.03.005

Daugaard, H. T., Cain, K., & Elbro, C. (2017). From words to text: Inference making mediates the role of vocabulary in children's reading comprehension. *Read and Writing*, *30*(8), 1773-1788. https://doi.org/10.1007/s11145-017-9752-2

De Jong, T. (2010). Cognitive load theory, educational research, and instructional design: Some food for thought. *Instructional Science, 38*, 105-134. doi:10.1007/s11251-009-9110-0

Denton, C. A., Tolar, T., Fletcher, J. M., Barth, A. E., Vaughn, S. R., & Francis, D. J. (2013). The effects of individualized Tier 3 reading intervention for students with persistent reading difficulties. *Journal of Educational Psychology*, *105*(3), 633-648. https://doi:10.1037/a0032581.

Dexter, D. D., & Hughes, C. A. (2011). Graphic organizers and students with learning disabilities: A meta-analysis. *Learning Disability Quarterly, 34*(1), 51-72. https://doi.org/10.1177/073194871103400104

Durlak, J. A. (2009). How to select, calculate, and interpret effect sizes. *Journal of Pediatric Psychology*, *34*(9), 917-928. https://doi.org/10.1093/jpepsy/jsp004

Elbro, C., & Buch-Iversen, I. (2013). Activation of background knowledge for inference making: Effects on reading comprehension. *Scientific Studies of Reading, 17*, 435-452. https://doi.org/10.1080/10888438.2013.774005

Fritschmann, N. S., Deshler, D. D., & Schumaker, J. B. (2007). The effects of instruction in an inference strategy on the reading comprehension skills of adolescents with disabilities. *Learning Disability Quarterly*, *30*(4), 245-262. https://doi.org/10.2307/25474637

Galy, E., Cariou, M., & Mélan, C. (2012). What is the relationship between mental workload factors and cognitive load types? *International Journal of Psychophysiology*, *83*(3), 269-275. https://doi.org/10.1016/j.ijpsycho.2011.09.023

Guthrie, J. T., & Davis, M. H. (2003) Motivating struggling readers in middle school through an engagement model of classroom practice. *Reading & Writing Quarterly*, *19*(1), 59-85, doi:10.1080/10573560308203

Hall, C., & Barnes, M. A. (2017). Inference instruction to support reading comprehension for elementary students with learning disabilities. *Intervention in School and Clinic*, *52*(5), 279-286. https://doi.org/10.1177/1053451216676799

Hart, S. G. (2006). NASA-Task Load Index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(9), 90-908. https://doi.org/10.1177/154193120605000909

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press. doi:10.1016/b978-0-08-057065-5.50010-5

Hogan, T. P., Bridges, M. S., Justice, L. M., & Cain, K. (2011). Increasing higher level language skills to improve reading comprehension. *Focus on Exceptional Children*, *44*(3), 1-19. https://doi.org/10.17161/fec.v44i3.6688

Hsieh, H-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*(9), 1277-1288. doi:10.1177/1049732305276687

Kaufman, A. S., & Kaufman, N.L. (2004). *Kaufman Brief Intelligence Test, Second Edition* (K-BIT-2). Pearson Assessment.

Kim, W., Linan-Thompson, S., & Misquitta, R. (2012). Critical factors in reading comprehension instruction for students with learning disabilities: A research synthesis. *Learning Disabilities Research & Practice, 27*(2), 66-78. https://doi.org/10.1111/j.1540-5826.2012.00352.x

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*(2), 163. https://doi.org/10.1037/0033-295X.95.2.163

Kucan, L., & Beck, I. (1997). Thinking aloud and reading comprehension research: Inquiry, instruction, and social interaction. *Review of Educational Research, 67*(3), 271-299. doi:10.2307/1170566

Laurie-Rose, C., Frey, M., Ennis, A., & Zamary, A. (2014). Measuring perceived mental workload in children. *The American Journal of Psychology, 127*(1), 107. doi:10.5406/amerjpsyc.127.1.0107

Leslie, L., & Caldwell, J.S. (2011). *Qualitative Reading Inventory-5*. Allyn & Bacon.

MacGinitie, W. H. (2000). *Gates-MacGinitie Reading Tests*. Riverside Publishing.

McLeskey, J., Barringer, M-D., Billingsley, B., Brownell, M., Jackson, D., Kennedy, M., Lewis, T., Maheady, L., Rodriguez, J., Scheeler, M. C., Winn, J., & Ziegler, D. (2017). *High-leverage practices in special education*. Council for Exceptional Children & CEEDAR Center. https://ceedar.education.ufl.edu/wp-content/uploads/2017/07/CEC-HLP-Web.pdf

Missouri Department of Education and Secondary Education. (2013). *MAP: LEA guide to the Missouri Assessment Program. Reading*. http://dese.mo.gov/college-career-readiness/asessment/assessment-technical-support-materials

Nachar, N. (2008). The Mann Whitney *U*: A test for assessing whether two independent samples come from the same distribution. *Tutorials in Quantitative Methods for Psychology, 4*(1), 13-20. https://doi.org/10.20982/tqmp.04.1.p013

O'Brien, E. J., Cook, A. E., & Lorca Jr., R. F. (2015). *Inferences during reading*. Cambridge University Press. https://doi.org/10.1017/CBO9781107279186

Paas, F., & van Merriënboer, J.J.G. (2020). Cognitive-load theory: Methods to manage working memory load in the learning of complex tasks. *Current Directions in Psychological Science*, *29*(4), 394-398. https://doi.org/10.1177/0963721420922183

Pressley, M. (2002). Metacognition and self-regulated comprehension. In A. E. Farstrup & S. J. Samuel (Eds.), *What research has to say about reading instruction* (3rd ed., pp. 291-309). International Reading Association. doi:10.1598/0872071774.13

Scammacca, N. K., Roberts, G. J., Cho, E., Williams, K. J., Roberts, G., Vaughn, S. R., & Carroll, M. (2016). A century of progress: Reading interventions for students in grades 4–12, 1914-2014. *Review of Educational Research*, *86*(3), 756-800. https://doi.org/10.3102/0034654316652942

Scammacca, N. K., Roberts, G., Vaughn, S., & Stuebing, K. K. (2015). A meta-analysis of interventions for struggling readers in grades 4-12: 1980-2011. *Journal of Learning Disabilities*, *48*(4), 369-390. https://doi.org/10.1177/0022219413504995

Scammacca, N., Roberts, G., Vaughn. S., Edmonds, M., Wexler, J., Reutebuch, C. K., & Torgesen, J. K. (2007). Interventions for adolescent struggling readers: A meta-analysis with implications for practice. RMC Research Corporation, Center on Instruction. https://www.centeroninstruction.org/files/Meta-analysis%20Struggling%20Readers1.pdf

Schrank, F. A., McGrew, K. S., & Woodcock, R. W. (2001). *Technical abstract* (Woodcock-Johnson III Assessment Service Bulletin No. 2). Riverside Publishing.

Stevenson, N. A., & Mussalow, P. R. (2018). The effects of planning, goal setting, and performance feedback on avoidance behaviors for struggling readers. *Journal of Positive Behavior Interventions, 21*(3), 171-180. doi:10.1177/1098300718804566

Swanson, E., Barnes, M., Fall, A.-M., & Roberts, G. (2017). Predictors of reading comprehension among struggling readers who exhibit differing levels of inattention and hyperactivity. *Reading & Writing Quarterly, 34*(2), 132-146. doi:10.1080/10573569.2017.1359712

Torgesen, J. K., Wagner, R., & Rashotte, C. (2012). *TOWRE–2 Test of Word Reading Efficiency*. Pro-Ed.

Whitehead, A. (1929). *The aims of education.* Macmillan.

Wolf, M. M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis*, *11*, 203-214. https://doi.org/10.1901/jaba.1978.11-203

Acknowledgments

The authors wish to acknowledge the invaluable contributions of the middle school staff and students in Laura Speed Elliot Middle School, Boonville School District, and the assistance of Principal Fred Smith.