Journal of Research in Reading

Volume 37, Issue 4
November 2014
Pages 356–374

Original Article

Gender differences in inference generation by fourth-grade students

Virginia Clinton , Ben Seipel, Paul van den Broek, Kristen L. McMaster, Panayiota Kendeou, Sarah E. Carlson, David N. Rapp

First published:
7 June 2012 Full publication history

DOI:
10.1111/j.1467-9817.2012.01531.x View/save citation

Funding Information

Abstract

The purpose of this study was to determine if there are gender differences among elementary school-aged students in regard to the inferences they generate during reading. Fourth-grade students (130 females; 126 males) completed think-aloud tasks while reading one practice and one experimental narrative text. Females generated a larger number and a greater proportion of reinstatement inferences than did males (Cohen's d = .34, p = .01; Cohen's d = .26, p = .04, respectively). In contrast, there was no evidence for gender differences in other types of think-aloud responses. These findings suggest that males and females differ in their use of cognitive processes that underlie reading comprehension, particularly with respect to the likelihood of retrieval of

information from episodic memory.

Enhanced Article Feedback

Implications for Practice

What is already known about this topic

- Females typically outperform males on reading comprehension assessments.
- Females typically outperform males on tasks requiring retrieval of information from episodic memory.
- Reinstatement inferences, which are important to reading comprehension, require retrieval of information from episodic memory.

What this paper adds

- Females generated more reinstatement inferences than did males in a think-aloud task.
- No other gender differences in think-aloud responses were found.

Implications for practice

- Reading interventions that facilitate inference generation may be particularly helpful for male students.

Gender differences in reading comprehension have been well documented. Females consistently outperform their male peers in most reading comprehension assessments at various stages of development and in many countries (DfES, 2007 ; Halpern, Benbow, Geary, Gur, Hyde & Gernsbacher, 2007 ; Logan & Johnston, 2009 ; Twist, Schagen & Hodgson, 2007). Moreover, boys are more likely than girls to have reading comprehension difficulties as indicated by a disconnect between a student's observed reading comprehension performance and predicted reading comprehension performance based on age and intellectual ability (Knopik, Alarcón & DeFries, 1998 ; Wheldall & Limbrick, 2010). This so-called 'gender gap' in reading comprehension may have profound consequences for males given the importance of literacy skills in both academics and employment (Riordan, 1999 ; Wood, 2003).

According to the *simple view of reading*, successful reading requires the use of two general sets of skills: decoding and linguistic comprehension (Gough & Tunmer, 1986). Reading for understanding requires the ability to identify printed words *and* the language comprehension skills necessary to derive meaning from those words. Results from studies on decoding skills as measured by nonsense word decoding have not revealed differences as a function of gender (Below, Skinner, Fearrington & Sorrell, 2010 ; Logan & Johnston, 2010). Therefore, gender differences in performance on reading comprehension assessments seem to be due to the comprehension activities rather than the decoding activities associated with reading. In this study, we investigate this line of reasoning by testing for gender differences in language comprehension skills. Specifically, we explore gender differences in cognitive processing while reading to expand on literature that has explored gender differences in reading attitudes and motivation as well as socialisation practices in which literacy-related activities are characterised as feminine (Gambell & Hunter, 2000 ; Logan & Johnston, 2009 ; Marinak & Gambrell, 2010).

Gender differences in cognition

The gender gap in performance on reading comprehension assessments is one specific instance that falls under the broader category of gender differences in cognition. Although there is no indication of a gender difference in general intelligence, considerable evidence suggests gender differences in performance on a variety of cognitive tasks (Halpern & LaMay, 2000 ; Johnson & Bouchard, 2007 ; Weiss, Kemmler, Deisenhammer, Fleischhacker & Delazer, 2003). To date, differences have been observed with cognitive tasks that focus on verbal, visual-spatial and quantitative skills (Hyde, 1990 ; Hyde & McKinley, 1997). Across studies, females typically performed better on verbal tasks and males typically performed better on visual-spatial and quantitative tasks (Kimura, 1999).

The patterns of gender differences in cognitive functioning, though, may be subtler than suggested by broad reference to verbal, visual-spatial and quantitative categories (Halpern, 2000 , 2004 ; Halpern & Wright, 1996). For example, there are some specific verbal tasks on which males typically outperform females (e.g., verbal analogies) and some specific visual-spatial and quantitative tasks on which females typically outperform males (e.g., memory for spatial location; Eals & Silverman, 1994 ; Gallagher, Levin & Cahalan, 2002). Thus, rather than grouping cognitive tasks into three broad categories (i.e., verbal, quantitative and visual-spatial), a more appropriate way of understanding such gender differences may be to consider the underlying processes that contribute to differential task performance (Halpern & Wright, 1996). One account that attempts to elucidate these underlying contributors is the *cognitive process taxonomy* (Halpern, 2000 , 2004), which reflects the cognitive processes necessary to perform tasks rather than the tasks themselves. The cognitive processes in which males and females differed were determined by examining the bulk of the empirical evidence in which gender differences in performance on cognitive tasks were found (Halpern, 2000 , 2004). Halpern (2000 , 2004) examined cognitive tasks that have obtained consistent findings with respect to gender differences, and then categorised the tasks based on the underlying cognitive processes needed to accomplish them. Based on this analysis, gender differences in cognitive tasks may be related to the retrieval of information from working memory and long-term memory. Working memory provides temporary storage for currently processed information before it is either discarded or encoded into long-term memory (Baddeley & Hitch, 1974). Long-term memory continually maintains and stores information in a relatively permanent form so it can be retrieved for later use (Atkinson & Shiffrin, 1968). The tasks in which males frequently outperform females involve maintaining and transforming mental representations in visual-spatial working memory, an aspect of working memory specific to visual and spatial information (Baddeley, 2010 ; Halpern 2000 , 2004). This enhanced retrieval from visual-spatial working memory is demonstrated with better performance on spatial perception, mental rotation and mechanical reasoning tasks (Maccoby & Jacklin, 1974). In contrast, the tasks in which females frequently outperform males involve retrieving information from long-term memory (Halpern, 2000 , 2004). This enhanced retrieval from long-term memory is demonstrated by better performance on tasks that involve verbal fluency, synonym generation and memory for spatial locations. However, results from several studies have indicated that gender differences in retrieval of information from long-term memory vary depending on the system of long-term memory involved (Halpern et al., 2007 ; Herlitz & Rehnman, 2008). In the next section, we discuss two systems of long-term memory and how they relate to reading comprehension, and any potential processing differences that might be ascribed to gender.

Long-term memory

Episodic memory and semantic memory are information processing systems in long-term memory (Tulving, 1972 , 2002). Episodic memory encodes information that one has personally experienced. Information in episodic memory is connected to the time and place that the information was acquired, such as remembering when and where one met his or her spouse or read a favourite book for the first time. Females typically outperform males in tasks involving retrieval of information from episodic memory, such as recognising faces (Guillem & Mograss, 2005 ; Lewin & Herlitz, 2002 ; Rehnman & Herlitz, 2007), recognising inkblots (Lewin, Wolgers & Herlitz, 2001), recalling lists of words (Kramer, Delis, Kaplan, O'Donnell & Prifitera, 1997) and remembering object locations (Vover, Postma, Brake & Imperato-McGinley, 2007). These tasks involve episodic memory because the stimuli (e.g., faces, word lists, etc.) were provided during the experimental session; therefore, the experimenters could determine that the participants had conscious recollections of their exposure to the stimuli (cf. Tulving, 1983).

In contrast, semantic memory encodes facts and knowledge about the world that one has learned, which are abstracted from the personal experience in which the knowledge was acquired (Tulving, 1972). Information in semantic memory is not connected to the time and place that the information was experienced, such as knowing the definitions of words and the capital cities of Europe without recalling when and where this information was learned. To date, there has been little evidence for gender differences in the retrieval of information from semantic memory (Halpern et al., 2007 ; Herlitz, Nilsson & Båchman, 1997). Therefore, the tendency for females to outperform males on retrieval of information from long-term memory outlined by the cognitive process taxonomy appears to be limited to information from episodic memory.

The observed gender differences specific to the retrieval of information from episodic memory provide a possible explanation for the gender differences in reading comprehension assessment scores. Recall that females have been found to outperform males in tasks involving retrieval of information from episodic memory. Readers use information from their representation of the text in episodic memory and their background knowledge in semantic memory to construct a coherent mental representation of the text (Van den Broek, Young, Tzeng & Linderholm, 1999), with such a representation necessary for successful reading comprehension (Van den Broek, Bohn-Gettler, Kendeou, Carlson & White, 2011). In order to construct a coherent mental representation of the text, one must produce inferences. Inference generation is necessary for coherence because texts are not written to explicitly state every relation between the ideas presented; a text that would do so would be cumbersome to read (Kintsch, 1998). One type of inference is the reinstatement inference in which information from previously read text is connected to currently read text (Singer, 1994). Reinstatement inferences require retrieval of information from episodic memory (McNamara & Magliano, 2009 ; Van den Broek & Gustafson, 1999). Therefore, if males and females differ in reinstatement inference generation because of retrieval of information from episodic memory, this difference may provide part of the reason for the observed gender difference in reading comprehension. We turn to the inferential processes that are associated with reading comprehension next.

Inference generation during reading

Broadly speaking, the inferences that can be made during reading generally fall into two broad categories: text connecting and knowledge based (Singer, 1994 ; Van den Broek, 1990). Text-connecting inferences connect two individual ideas presented in the text (Kintsch, 1998), and can be further categorised into connective and reinstatement inferences (Linderholm & van den Broek, 2002). Connective inferences connect the currently read sentence with information from the sentence immediately preceding it. Reinstatement inferences connect the currently read sentence with information from sentences occurring in the text earlier than the immediately preceding sentence. Information previously stated in the text used to generate text-connecting inferences is likely retrieved from working memory (for connective inferences) or episodic memory (for reinstatement inferences; Trabasso & Magliano, 1996 ; Van den Broek, 1994). In contrast to text-connecting inferences, knowledge-based inferences are extratextual in that they involve information from outside the text (Graesser, Singer & Trabasso, 1994). These knowledge-based inferences are generated by the reader to predict or explain a text (Graesser, Millis & Zwaan, 1997). The information from background knowledge that is used to generate knowledge-based inferences is retrieved from semantic memory (Kintsch & van Dijk, 1978).

The present study

The research question addressed in this study is whether females generate more reinstatement inferences than males do. If so, this difference in the process of reading would provide at least a partial albeit intriguing explanation for the gender differences observed on comprehension assessments. Because reinstatement inference generation requires retrieval of information from episodic memory (Graesser et al., 1994) and because females have been found to have an advantage over males in such retrieval (Herlitz et al., 1997), we hypothesise that females will generate more reinstatement inferences than do males. Because connective inferences require retrieval of information from working memory (Van den Broek, 1994), we expect to observe no gender difference in connective inference generation. Because knowledge-based inferences require retrieval of information from background knowledge from semantic memory (Kintsch & van Dijk, 1978), we expect to observe no gender difference in knowledge-based inference generation.

Inference generation is relevant to pedagogy because of its necessity for successful reading comprehension (Cain & Oakhill, 2007 ; Graesser et al., 1997 ; Van den Broek, 1990). For example, text-connecting inferences, both connective and reinstatement, are necessary to enable the reader to monitor situations and actors in the text (Graesser et al., **1994**). The pedagogical implications of inference generation are particularly salient in the later elementary school grades. Around the fourth grade, readers transition from *learning to read* primarily to *reading to learn* primarily (Best, Floyd & McNamara, **2008**). Most readers have mastered adequate decoding skills and, around the fourth grade, now must draw on their language comprehension skills for the acquisition of new material from expository sources such as science and social studies textbooks (Best et al., **2008**). Without effective inference generation, students in later school grades struggle with reading to learn; they exhibit difficulties in understanding the material presented to them in texts (Best, Rowe, Ozuru & McNamara, **2005** ; Cain, Oakhill & Bryant, **2004**). For these reasons, fourth-grade students were the population of interest in this study.

To examine inference generation during reading, we employed a think-aloud task (Ericsson & Simon, **1993** ; Pressley & Afflerbach, **1995**). Think-aloud tasks ask individuals to read a unit of text (e.g., a sentence), and state what they are thinking about while they are reading. This procedure allows researchers to examine the process of reading as it unfolds (Coté & Goldman, **2004** ; Van den Broek, Lorch, Linderholm & Gustafson, **2001**). Think-aloud protocols can be particularly useful in examining inference generation because they capture the inferential process as it happens and provide insight into how individuals attempt to build mental representations of the texts they are reading (Ericsson & Simon, **1993** ; Pressley & Afflerbach, **1995**). The responses provided in the think-aloud protocols allowed us to address the question of whether gender differences might emerge with respect to inference generation.

Method

Participants

The participants in this study were 256 fourth-grade students (*M* = 9.44 years, *SD* = 0.94 years; 130 females, 126 males). These data had been collected as part of a study focused on reading comprehension instruction. The students had not been exposed to any new reading comprehension instruction prior to the data being collected. Students were from 13 classrooms in three elementary schools in two school districts of a large, upper Midwestern urban area in the United States. The first district, from which one school participated, had a total population of 4,540 students; 27% of the students received free or reduced-price lunch, 11% qualified for special education services, and 6% qualified for English Language Learner (ELL) services. The second district, from which two schools participated, had a total population of 9,063 students; 29% of the students received free or reduced-price lunch, 12% qualified for special education services, and 10% qualified for ELL services. Students with developmental or intellectual disabilities were not included in the study to avoid confounding reading comprehension difficulties with issues related to general cognitive processing. ELLs were not included in the study to avoid confounding reading comprehension difficulties with issues related to non-native English proficiency. The racial background of the majority of the participants was primarily Caucasian (75.2%), with other participants identifying as African American (13.4%), Asian (2.4%), Hispanic (4.5%) or not providing their racial background (2.4%). Participants had a wide range of reading comprehension skills based on the Gates–MacGinitie Reading Comprehension subtest scores (percentile ranks between 1st and 99th percentile; *M* = 50.70, *SD* = 21.00; MacGinitie, MacGinitie, Maria & Dreyer, **2000** ; see Measures) and teacher reports. Students were compensated with a $10 gift card to a major retailer.

Measures and materials

Participants completed the Gates–MacGinitie Reading Test (GMRT) Comprehension subtest (Fourth Edition, Level 4, Forms S and T; MacGinitie et al., **2000**) as part of a screening battery used in the larger study. Research assistants group-administered the GMRT reading comprehension subtest to the participants in their classrooms. The subtest is a nationally normed assessment consisting of 11 short (fewer than 20 sentences) texts with two to six multiple-choice questions after each text. Participants have 35 minutes to read six narrative texts, four expository texts and one setting text (i.e., a description that did not include a temporal-driven plot) and answer 48 multiple-choice questions. The maximum score is 48. The reported Kuder–Richardson 20 reliability with a fourth-grade student sample is 0.93 (MacGinitie, MacGinitie, Maria & Dreyer, **2002**). Normal curve equivalents (NCE) from the Gates–MacGinitie scores were used in all analyses.

One practice and one experimental text were used to collect data for the think-aloud task (see Appendix Appendix for practice and experimental texts). The experimental text was 21 sentences long containing 202 words and 809 characters with a Flesch Reading Ease of 85.6 and a Flesch–Kincaid Grade Level of 3.5.

Procedure

Students were administered the think-aloud task during individual sessions. The think-aloud task was used in this study to provide an account of the cognitive processes in which readers engage during reading (cf. Pressley & Afflerbach, **1995**). Students read one practice and one experimental text. Each sentence from each of the texts was printed on an index card. Students read each sentence at their own pace. Before beginning the think-aloud task, the experimenter modelled a variety of responses and explained to the student that they would answer two comprehension questions at the end of each text (see Appendix Appendix for script). Students then read each sentence aloud to the experimenter. After reading each sentence aloud, students verbally reflected on the sentence (e.g., Pressley & Afflerbach, **1995** ; Van den Broek et al., **2001**). Students were not permitted to look back to previously read cards. If a student did not comment on a sentence, the experimenter encouraged the participant to provide a response with nonleading prompts such as, 'What are you thinking now?' or 'Please keep talking'. After each student completed a story, the experimenter asked the student two comprehension questions, which could be answered with a 'yes' or 'no'. Think-aloud responses were recorded on audiocassette tapes and transcribed for analysis.

Think-aloud coding

Verbal data from the think-aloud procedure were transcribed. Each of the idea units (noun–verb combinations that express a single idea) from participant responses was coded as one of the following categories: connective inferences (connect information in the sentence immediately preceding the currently read sentence [from working memory] to the currently read text), reinstatement inferences (connect information from the previously read text prior to the sentence immediately preceding the currently read text [from episodic memory] to explain the currently read text), knowledge-based inferences (invoke background knowledge [from semantic memory] to explain the currently read text or develop predictions for the text), associations (elaborations irrelevant to the text), paraphrases (summary of the currently read sentence), text repetitions (verbatim reiterations of the text), questions (questions about the text), affective responses (emotional reactions), evaluative comments (opinions about the text), meta-cognitive comments (reflections of the reader's own understanding or lack of understanding; think-aloud coding based on Kendeou & van den Broek, 2007 ; Van den Broek et al., 2001). Inferences were further coded as valid if they were accurate or consistent with the text, or invalid if they were inaccurate or inconsistent with the text. Only valid inferences were included in analyses. See Table 1 for examples of think-aloud responses for each code. Inter-rater reliability for coding was good, K = .89. Disagreements between coders were resolved through discussion.

Table 1. Examples of think-aloud responses

Code	Definition	Original text	Sample response
Connective inference	Connection of the currently read sentence with the immediately preceding sentence	Brian decided to put on a magic act. He bought some fancy magic cards and a hat.	'Brian got the cards and hat for his magic act'.
Reinstatement inference	Connection of the currently read text with the previously read text prior to the immediately preceding sentence	The tricks were hard to do [one intervening sentence]. He asked a magician if he would teach him.	'Brian probably did this because the tricks were hard to do'.
Knowledge-based inference	Explanations or predictions of the text based on background knowledge	Brian learned how to perform different kinds of magic.	'I bet Brian learned how to pull a rabbit out of hat!'
Invalid inference	Inferences that are inaccurate or inconsistent with the text	Brian decided to put on a magic act.	'He's probably going to go to the playground'.
Associations	Elaborations irrelevant to the text	Brian put on a magic act.	'I saw a magician once'.
Paraphrase	Summary of the text sentence	He selected tricks he could do best.	'He chose the tricks he was really good at'.
Text repetition	Verbatim reiterations of the text	Brian visited every day for a month.	'Brian visited every day for a month'.
Questions	Questions about the text	He bought some fancy magic cards and a hat.	'What kind of cards did he get?'
Affective response	Emotional reactions	He made some pretty flowers come out of her hat and then made the	'Wow!'

		flowers disappear.	
Evaluative comments	Opinions about the text	Brian put together a good act, showing his teacher.	'This story's interesting'.
Meta-cognitive comments	Reflections of the readers' own understanding or lack of understanding	He watched how each routine was done and practised a lot.	'I know what a routine is'.

Dependent variables

The purpose of this study was to determine if there were gender differences in students' generation of reinstatement inferences. To do this, we examined the number of reinstatement inferences generated and the proportion of reinstatement inferences of all responses provided by participants during the think-aloud task. The number of reinstatement inferences was examined so that production of reinstatement inferences could be compared across genders. The proportion of reinstatement inferences was examined so that the entirety of participant responses would be considered (Ericsson & Simon, **1993**). These proportions were calculated by dividing the number of knowledge-based and reinstatement inferences by the total number of think-aloud responses produced. This calculation allowed for an examination of inference generation in the context of the other types of responses produced during text processing (Trabasso & Magliano, **1996**). Because of potential gender differences in production rates (Swann, **1998**), analysis of the raw idea units might inaccurately depict patterns of generation, based solely on students of one gender speaking more frequently rather than on the quality of think-aloud responses.

Generation of other processes for think-aloud responses was examined in a manner parallel to that used for the reinstatement inferences, namely by calculating the number and the proportion of each of the codes generated. This additional analysis allowed us to determine if any gender differences revealed were specific to reinstatement inference generation or to think-aloud responses in general. To facilitate this analysis, certain codes of think-aloud responses were combined to provide a profile of the effects. Invalid inferences and associations were combined because they both involve information tangential or irrelevant to the text. Affective responses and opinions were combined because they both involve emotional reactions to the text. Meta-cognitive comments and questions were combined because they both involve readers either commenting on their own understanding or their lack of understanding. Paraphrases and text repetitions were combined because they both involve responses constrained to the currently read sentence. In addition, gender differences in the total number of think-aloud responses produced were analysed to determine if one gender produced more responses overall than another.

Results

The means and standard deviations for the number and proportions of each code of the think-aloud responses by gender are provided in Tables 2 and 3, respectively.

Table 2. Means and standard deviations of the number of think-aloud responses for each code by gender

	Males	Females	All
Connective	1.46 (1.64)	1.66 (1.49)	1.56 (1.57)
Reinstatement	2.84 (2.81)	3.90 (3.31)	3.38 (3.12)
Knowledge based	14.86 (7.26)	15.63 (8.63)	15.25 (7.98)
Invalid/associations	1.36 (3.36)	1.40 (4.18)	1.38 (3.79)
Paraphrase/TR	7.56 (6.29)	8.28 (6.02)	7.93 (6.15)
MC/questions	0.80 (1.57)	0.88 (2.61)	0.84 (2.16)
Opinions/affective	0.91 (2.09)	1.02 (1.94)	0.96 (2.01)
Total	29.95 (10.93)	32.92 (11.39)	31.46 (11.24<)

Notes: *N* = 256 (130 females; 126 males). Connective = connective inferences, Reinstatement = reinstatement inferences, Knowledge based = knowledge-based inferences, Invalid/associations = invalid inferences and associations, Paraphrase/TR = paraphrases and text repetitions, MC/questions = meta-cognitive comments and questions, Opinions/affective = evaluative comments and affective responses, Total = total number of responses.

Table 3. Means and standard deviations of the proportion of think-aloud responses for each code by gender

	Males M (*SD*)	Females M (*SD*)	All M (*SD*)
Connective	0.05 (0.05)	0.05 (0.05)	0.05 (0.05)
Reinstatement	0.09 (0.08)	0.11 (0.09)	0.10 (0.08)
Knowledge based	0.51 (0.20)	0.48 (0.20)	0.49 (0.20)
Invalid/associations	0.04 (0.08)	0.04 (0.07)	0.04 (0.08)
Paraphrase/TR	0.24 (0.19)	0.25 (0.19)	0.25 (0.19)
MC/questions	0.03 (0.07)	0.03 (0.10)	0.03 (0.08)
Opinions/affective	0.03 (0.06)	0.03 (0.07)	0.03 (0.07)

Notes: *N* = 256 (130 females; 126 males). Connective = connective inferences, Reinstatement = reinstatement inferences, Knowledge based = knowledge-based inferences, Invalid/associations = invalid inferences and associations, Paraphrase/TR = paraphrases and text repetitions, MC/questions = meta-cognitive comments and questions, Opinions/affective = evaluative comments and affective responses.

Parametric statistics

One-way ANOVAs were conducted with the number of think-aloud responses and proportion of think-aloud responses for each process as the dependent variable and gender as the independent variable. We hypothesised that females would generate more reinstatement inferences than would males, which was supported with respect to the number of think-aloud responses for each code, $F(1, 254) = 7.58$, $p = .01$, Cohen's $d = .34$. No other gender differences in the number of responses for other think-aloud codes were found. However, a gender difference in the overall total number of think-aloud responses was found with females producing more responses than did males, $F(1, 254) = 4.53$, $p = .03$, Cohen's $d = .31$. Regarding the proportions of think-aloud responses, females also generated a greater proportion of reinstatement inferences than did males, $F(1, 254) = 4.36$ $p = .04$, Cohen's $d = .26$, which is consistent with the above hypothesis. No gender differences in the proportions of other processes in think-aloud responses were found.

Nonparametric statistics

Because of non-normality in the distribution of several of the think-aloud responses (based on skewness as indicated by means and standard deviations), nonparametric statistics (Independent Mann–Whitney U tests) were conducted. Regarding the number of think-aloud responses, the distributions of reinstatement inferences for the two genders were found to be different, $U = 6{,}627$, $z = -2.66$, $p = .01$, as is consistent with the findings from the parametric statistics. In addition, a gender difference in the overall number of think-aloud responses was found with females producing more responses than did males, $U = 6{,}768$, $z = -2.40$, $p = .02$, as is consistent with the findings from the parametric statistics. The distributions of the number of all other think-aloud responses were not found to be different. Regarding the proportions, the distributions of reinstatement inferences for the two genders were found to be different, $U = 7{,}047.5$, $z = -1.93$, $p = .05$, as is consistent with the findings from the parametric statistics. The distributions of the proportions of all other think-aloud responses were not found to be different.

Reading comprehension skill

We examined the possibility that the observed gender differences in reinstatement inference generation reflected differences in basic reading comprehension skill, given that positive associations between inference generation and performance on reading comprehension assessments have been reported in previous research (Cain et al., 2004 ; Cromley & Azevedo, 2007). However, we did not find a significant correlation between GMRT NCE and the number of reinstatement inferences, $r = .09$, $p = .14$, or the proportion of reinstatement inferences, $r = .04$, $p = .5$. We tested for gender differences with a one-way ANOVA with gender as the between-subjects variable and GMRT NCE as the dependent variable. We did not find gender differences in GMRT NCE, $F(1, 254) = 0.25$, $p = .62$,

(*M* = 20.02; *SD* = 22.09 for males; *M* = 51.35; *SD* = 19.96 for females).

Discussion

The purpose of this study was to examine potential gender differences in reinstatement inference generation. To address this purpose, we examined think-aloud responses produced by fourth-grade students reading a narrative text. We hypothesised that females would generate more reinstatement inferences than would males given previous findings on gender differences in retrieval of information from episodic memory. The findings supported the hypothesis: females generated a greater number and a larger proportion of reinstatement inferences than did males. In contrast, there were no gender differences in connective inference generation, which involves retrieval of information from working memory, nor were there gender differences in knowledge-based inference generation, which involves retrieval of information from semantic memory.

For both males and females, knowledge-based inferences and paraphrases or repetitions of the currently read sentence were the most common think-aloud responses. In addition, certain think-aloud responses, such as invalid inferences and associations, were uncommon for both genders. There were no gender differences in think-aloud responses other than reinstatement inferences. However, there are neither theoretical bases nor previously established empirical findings that would lead one to expect differences in think-aloud responses other than reinstatement inferences. There were gender differences in the overall production of idea units with females producing more responses while thinking aloud than did males. This is likely because females generated significantly more reinstatement inferences than did males. However, slight and not significant differences in other think-aloud responses that females produced more than males (e.g., opinions and affective responses) likely also contributed to the gender difference in overall production of idea units.

The finding that females generated more reinstatement inferences than did males is consistent with both the cognitive process taxonomy and previous findings on gender differences in retrieval of information from episodic memory (Halpern, **2004**). Generally, females have been found to perform better than males when retrieving information from long-term memory, as is categorised by the cognitive process taxonomy (Halpern, **2000**). The prior knowledge necessary to generate reinstatement inferences must be retrieved from long-term memory (Graesser et al., **1994**). The fact that we observed no gender difference with respect to connective inferences also is consistent with this account as connective inferences require retrieval from working memory, not from long-term memory (Halpern, **2000** ; Van den Broek, **1990**). However, results from previous studies indicate that gender differences in retrieval of information from long-term memory are specific to episodic memory (Guillem & Mograss, **2005** ; Lewin et al., **2001** ; Lewin & Herlitz, **2002**). The findings from this study are consistent with previous findings with gender differences regarding episodic memory and semantic memory. Because the students in this study were not permitted to look back to the previously read sentences, information from the previously read text used to generate reinstatement inferences was likely retrieved from episodic memory. In contrast to episodic memory, there are no reported findings of gender differences in retrieving information from semantic memory. This lack of findings regarding semantic memory is consistent with the results from our study in which there were no gender differences in knowledge-based inference generation, which necessarily involves retrieval of information from semantic memory.

In this study, we did not find gender differences in performance on the reading comprehension assessment. This was surprising considering the previously discussed findings in which females outperformed males on reading comprehension assessments. There are a few possible reasons for this result. One is that our sample size may not have been large enough to detect a gender difference in reading comprehension performance. Several of the studies in which gender differences in reading comprehension assessment were found had sample sizes of thousands of participants (DfES, 2007 ; Halpern et al., 2007 ; Twist et al., 2007). Another reason may be that male readers in our study who generated fewer reinstatement inferences may have been able to compensate for a lack of inference generation with reading skills for which we did not assess, such as word identification or vocabulary knowledge. Although previous studies have not indicated a gender difference in word identification (Below et al., 2010 ; Logan & Johnston, 2010) or vocabulary knowledge (Hyde, 2005), such differences may have existed in our sample. Finally, the reading comprehension assessment we used (GMRT) has not been previously indicated to have gender differences in test scores nor were its test scores found to be positively associated with reinstatement inference generation. It is possible that reading comprehension assessments used in studies in which gender differences were found were more dependent upon reinstatement inference generation than the GMRT. Given that different reading comprehension tests assess different components of reading (Keenan, Betjemann & Olson, 2008), it is possible that another reading comprehension test would have revealed gender differences in test scores in our sample. Future investigations with larger sample sizes and broader, more holistic reading assessments than used in our study may be warranted to fully understand the role, if any, of reinstatement inference generation in gender differences in performance on reading comprehension assessments.

A future investigation with more comprehensive assessments of reading than we used in our study could find that the gender differences in reinstatement inference generation may partially explain the gender difference in performance on reading comprehension assessments. Reinstatement inferences are necessary to produce a coherent mental representation of a text (Graesser, Bertus & Magliano, 1995). Readers who produce fewer reinstatement inferences would have difficulty constructing the mental representations that are associated with adequate comprehension. Therefore, readers who generate fewer reinstatement inferences would logically perform more poorly on assessments of reading comprehension than their peers who generate more reinstatement inferences.

Implications

Gender differences in cognitive processes are changeable; they are by no means an inherent or necessary aspect of being male or female (Halpern, 2004). Previous studies have found that with practice and guidance, people can improve on tasks typically found to reveal gender differences in performance, regardless of gender (Law, Pellegrino & Hunt, 1993 ; Subrahmanyam & Greenfield, 1994). Therefore, the findings from this study may be used to inform the development of interventions. For example, McNamara (2004) reports that teaching readers to connect information within the text, as is done with reinstatement inferences, can improve reading comprehension. Therefore, interventions that encourage the generation of reinstatement inferences by teaching students to connect the information in the text, such as SERT and iSTART (McNamara, 2004 ;

McNamara, O'Reilly, Rowe, Boonthum & Levinstein, 2007), may be particularly useful for helping male (and female) students who struggle with reading comprehension.

It is important to note that this study's findings were with fourth-grade students whose reading comprehension skills are still in development. It is possible that gender differences in reinstatement inference generation may not be observed with readers at other stages of development. A future longitudinal or cross-sectional study could address the question of whether these gender differences in reinstatement inference generation persist across the lifespan.

This study raises the question as to why gender differences in cognitive processing (i.e., retrieval of information from episodic memory) might exist. We did not seek to answer this question in our study. However, researchers have previously suggested environmental explanations to account for gender differences in language and retrieval of information from episodic memory. One explanation for gender differences in language use, the *gender-as-culture hypothesis* , is that males and females typically socialise with members of their own gender, thus creating gender-specific subcultures (Maltz & Borker, 1982). Because of these gender-specific subcultures, males and females develop different patterns of language use (Mulac, Bradac & Gibbons, 2001). Another explanation based on environmental influences for gender differences in retrieval of information from episodic memory is that there may be different expectations placed on children of different genders in terms of remembering information (Herlitz et al., 1997). It may be useful for future researchers to further investigate the source of gender differences in cognitive processing, in the hopes of potentially ameliorating these differences through intervention-based activities (cf. Mulac, 1998 ; Halpern, 2000 for more information on potential sources of gender differences).

Conclusion

Gender differences have long been a central consideration of researchers in educational and psychological domains. Contemporary accounts of gender differences have focused on the ways in which underlying cognitive processes might describe or explain the nature of any observed differences (Halpern, 2004). Beyond these basic concerns, similar attention to gender differences has been of interest to researchers studying individual differences in academic settings. This study provides a deeper understanding of gender differences in reading comprehension performance and ideas for future interventions to help struggling male readers. The current study indicates that gender differences in reinstatement inferences may be a function of differences in readers' propensities or abilities for retrieval of information from episodic memory. An exciting possibility for future investigation involves directly encouraging such retrieval as readers process texts as a potential means of reducing gender differences and generally enhancing reading comprehension.

Acknowledgements

This data set was originally collected as part of a study supported with funding from the Institute of Education Sciences (Grant R305G040021). This study was made possible through funding from the US Department of Education (USDE) Award #R305C050059 (Minnesota Interdisciplinary Training in

Education Research [MITER] Program). We wish to thank the students and teachers who participated in the study and the many undergraduate and graduate research assistants who provided classroom support, developed materials, and collected and coded data. We also thank Dr Mark Davison for his helpful comments on this project.

Appendix A

Practice text

Once there was a boy named Jimmy.

One day, Jimmy saw his friend Tom riding a new bike.

He wanted to buy a bike.

Jimmy spoke to his mother.

Jimmy's mother refused to get a bike for him. He was very sad.

The next day, Jimmy's mother told him that he should have his own savings.

Jimmy wanted to earn some money.

He asked for a job at a nearby grocery store.

He made deliveries for the grocery store.

He earned a lot of money.

Jimmy went to the department store.

He walked to the second floor.

Jimmy bought a new bike.

Experimental text

Brian's Magical Skill

Brian liked to perform for people.

His teacher said that she was going to choose someone to be in a school show that would be in the spring.

Brian wanted to be chosen.

He sang a song for his teacher.

It was not very good.

His teacher did not choose him.

Brian decided to put on a magic act.

He bought some fancy magic cards and a hat.

He tried to do some tricks on a table.

The tricks were difficult to do.

Brian wanted to learn other kinds of activities.

He asked a magician if he would teach him.

The magician agreed to teach him.

Brian visited every day for a month.

He watched how each routine was done and practiced a lot.

Brian learned how to perform different kinds of magic.

He selected tricks that he could do best.

He put together a good act, showing his teacher.

He made some pretty flowers come out of her ear and then made the flowers disappear.

The magic act was a hit and was selected for the show.

Appendix B

Experimenter script for think-aloud procedure

Now, you are going to read a story, one sentence at a time.

Each time you read a sentence, you will tell me what you are thinking.

When you are done reading, I will ask you a couple of questions about the story.

Here is a story we will use for practice.

Each sentence of the story is on a separate card.

Each time you read a sentence, you will talk to me about what you are thinking as you read the sentence.

After you are done reading and talking, I will ask you two questions about the story and you will answer either ‘Yes’ or ‘No’.

I will record your comments and answers using this tape recorder. OK?

First, I will give you an example of how to talk aloud.

I will read each line out loud and talk about what I am thinking.

I will go through the first half of this story to show you, and then you will have a chance to try it. OK?

‘Jimmy and the new Bike’

OK, this is going to be a story about Jimmy and a bike. Maybe he's going to be in a bike race?

‘Once there was a boy named Jimmy’.

OK, Jimmy is the main character in the story.

‘One day, Jimmy saw his friend Tom riding a new bike’.

Jimmy saw his friend with a bike.

'He wanted to buy a bike'.

Jimmy wants a new bike like his friend Tom so maybe they can ride together. Maybe they are both going to be in a race?

'Jimmy spoke to his mother'.

He's going to ask his mother for a new bike like Tom's.

'Jimmy's mother refused to get a bike for him'.

OK, mom said no, so Jimmy will have to come up with another idea for getting a bike.

Now it's your turn.

Read each sentence out loud and tell me what you are thinking – even if it seems obvious.

Make sure you understand what you read because you cannot go back to cards once you have talked about them.

Biographies

Virginia Clinton is a postdoctoral research associate for the National Center for Cognition and Mathematics Instruction at the University of Wisconsin – Madison. Her research interests are in the cognitive processes involved in text comprehension, specifically inference generation.

Ben Seipel, PhD, is an Assistant Professor in the School of Education at California State University, Chico, where he primarily teaches courses in Educational Psychology and assessment to prepare future teachers for the classroom. His research interests broadly include reading comprehension and vocabulary acquisition.

Paul van den Broek is Professor in Education and Child Studies and Director of the Brain and Education Lab at the University of Leiden (the Netherlands) and in Cognitive Sciences at the University of Minnesota. His expertise is on cognitive processes in reading comprehension, and on the application to reading comprehension interventions. He is lead investigator on a 7-year longitudinal investigation of the development of reading and pre-reading skills, and on IES-funded studies on the cognitive processes by readers and on the development of school-based interventions. He has published more than 120 journal articles and chapters and has co-edited several books.

Kristen L. McMaster, PhD, is an Associate Professor of Special Education in the Department of Educational Psychology, University of Minnesota. She received her PhD in special education from Vanderbilt University. Her research interests include creating conditions for successful response to intervention of students at risk and students with disabilities. Specific research focuses on (1) promoting teachers' use of data-based decision-making and evidence-based instruction and (2) developing individualised interventions for students for whom generally effective instruction is not sufficient.

Panayiota Kendeou is an Assistant Professor of Educational Psychology at Neapolis University Pafos, Cyprus. Her current research focuses on the cognitive processes that support memory and learning in the context of reading comprehension. She currently serves as an Associate Editor of the *Journal of Research in Reading* and on the editorial boards of *Scientific Studies of Reading* , *Contemporary Educational Psychology* , *Learning and Instruction* , and *Reading Psychology* .

Sarah E. Carlson, PhD, is an Institute of Education Sciences (IES) postdoctoral research fellow at the Center of Teaching and Learning (CTL) at the University of Oregon. She holds a doctorate and master's degree in educational psychology within the psychological foundations of education, learning and cognition programme. Her current research is focused on reading comprehension, assessment, and interventions for struggling elementary and adolescent readers.

David N. Rapp is Associate Professor at Northwestern University jointly appointed in the School of Education and Social Policy, and in the Department of Psychology. His research focuses on the cognitive mechanisms underlying successful and unsuccessful learning. He is currently Associate Editor of the *Journal of Educational Psychology and Discourse Processes* .

References