Looking back: reasoning and metacognition with narrative texts

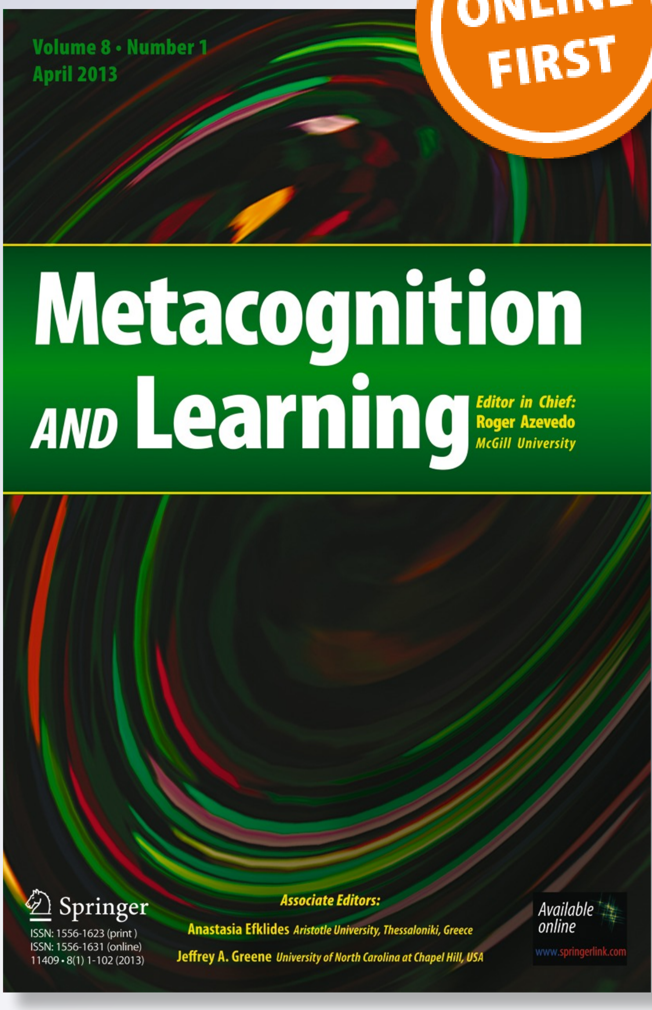
Bridget A. Franks, David J. Therriault, Miriam I. Buhr, Evelyn S. Chiang, Claire M. Gonzalez, Heekyung K. Kwon, Jenni L. Schelble, et al.

Metacognition and Learning

ISSN 1556-1623

Metacognition Learning
DOI 10.1007/s11409-013-9099-2

ONLINE FIRST



Metacognition Learning
DOI 10.1007/s11409-013-9099-2

Looking back: reasoning and metacognition with narrative texts

Bridget A. Franks · David J. Therriault · Miriam I. Buhr · Evelyn S. Chiang · Claire M. Gonzalez · Heekyung K. Kwon · Jenni L. Schelble · Xuesong Wang

Received: 1 July 2010 / Accepted: 16 April 2013

Abstract This study explored the abilities of 5th, 8th, and 10th graders, and College students to reason logically about what they read. Both students' metacognitive behavior (looking back at previously read text) and their performance on logical deduction questions were recorded and analyzed in a reading task. Conditional logic premises and deductive questions were embedded in three narratives containing premise information that was factual (True Story), contrary to fact (False Story), or unverifiable via common world knowledge (Neutral Story). The texts and questions were presented one sentence at a time on a computer screen; participants controlled the presentation of sentences. For answering the questions, three response tasks were devised. One task (labeled Generate) required readers to generate their own logical conclusions in response to

B. A. Franks (✉)
Teacher Education Department, 406M Roskens Hall, College of Education,
University of Nebraska at Omaha, 6005 Dodge St, Omaha, NE 68182-0163, USA
e-mail: bafranks@unomaha.edu

D. J. Therriault · M. I. Buhr
School of Human Development and Organizational Studies in Education
University of Florida, Gainesville, FL, USA

H. K. Kwon
Seoul National University, Seoul, South Korea

Present Address:
E. S. Chiang
Department of Psychology, University of North Carolina-Asheville, Asheville, NC, USA

Present Address:
C. M. Gonzalez
Department of Foundations and Secondary Education
University of North Florida, Jacksonville, FL, USA

Present Address:
J. L. Schelble
Department of Psychology, Western Kentucky University, Bowling Green, KY, USA

Present Address:
X. Wang
Department of Curriculum and Pedagogy, Texas A&M International University, Laredo, TX, USA

Published online: 18 May 2013

deduction questions. Two tasks (labeled Valid and Invalid) required readers to evaluate logically valid or logically invalid conclusions drawn by story characters in the texts. Students in early and late adolescence looked back more when asked to evaluate logical conclusions than when asked to generate conclusions on their own; College students' lookback frequencies were not significantly affected by response task, but were greater overall than those of younger students. With conditional forms requiring an uncertainty response (Affirmed Consequent and Denied Antecedent), readers looked back more when evaluating logically invalid conclusions than when evaluating logically valid ones. Readers of all ages were more likely to agree with story characters' (valid) uncertain conclusions with the AC and DA forms than they were to disagree with story characters' (invalid) certain conclusions to these forms. Both lookback frequency and performance on logic questions were lowest when readers were required to reason from contrary to fact premises.

Keywords Reasoning · Text comprehension · Metacognition · Logic · Adolescent development

The time we know as adolescence encompasses some of the most interesting cognitive changes of the human lifespan. At the beginning of the second decade of life, children can be logical and systematic in some forms of problem-solving, but still find it difficult to reflect on their own thought processes and monitor their own comprehension (Deloache et al. 1998). By the end of this decade, many have emerged as abstract reasoners (Inhelder and Piaget 1958) whose mental affinity for possibilities allows them to generate and test hypotheses, analyze symbolism in literature, and write program codes for computer applications. The developing reasoning skills that make this transformation possible are needed for the demanding cognitive tasks expected of adolescents in middle and high school, which include the comprehension and critical evaluation of increasingly complex texts. Secondary level reading is increasingly about thinking; readers must understand logical arguments, test hypotheses, and evaluate claims made by authors on the basis of evidence and convincing reasons.

Models of logical reasoning development offer useful predictions about the metacognition involved in reasoning tasks; in this study, we explored those predictions in a reading context. We observed the metacognitive behavior (looking back at previously read text) of students in grades 5, 8, 10, and College in response to questions requiring logical conclusions about text information. We examined the effects of either generating their own conclusions or evaluating conclusions made by story characters on students' lookback frequencies and their logical conclusions.

Conditional reasoning

Logical reasoning involves determining what would follow from stated premises if they were true. One widely studied form of logical reasoning is the conditional, which has the major premise "*if p*, *then q*." Standard logic offers four argument forms based on this premise:

Modus Ponens (**MP**): If *p* then *q*; *p* is true; therefore, *q* is true.

Example: *If this is a farm*, *then there is a tractor*. *It is a farm*. *Therefore*, *there is a tractor*.

Modus Tollens (**MT**): If *p* then *q*; *q* is not true; therefore, *p* is not true.

Example: *If this is a farm*, *then there is a tractor*. *There is not a tractor*. *Therefore*, *this is not a farm*.

Affirmation of the Consequent (**AC**): If *p* then *q*; *q* is true (undecidable).

Example: *If this is farm*, *then there is a tractor*. *There is a tractor*. *It is undecidable whether or not this is a farm*.

Denial of the Antecedent (DA): If *p* then *q*; *p* is not true (undecidable).

Example: *If this is a farm, then there is a tractor. This is not a farm. It is undecidable whether or not there is a tractor.*

MP and MT are commonly described as certain (or determinate) logical forms, since, according to standard logic, their conclusions can be determined with certainty. AC and DA are described as uncertain (or indeterminate) forms, because (again according to standard logic) conclusions cannot, with certainty, be drawn from them. Natural language interpretations of conditionals, however, often lead people to interpret AC and DA as certain forms. As noted by Daniel and Klaczynski (2006), such *invited* conclusions are intuitively appealing and often accurate in everyday language. For example, the AC conditional, "If you lift weights regularly, you will have strong muscles," with the minor premise, "Harriet has strong muscles," invite the conclusion that Harriet lifts weights regularly or that her doing so is very likely. This interpretation of a conditional may be activated automatically, and results in a certain conclusion; to draw a conclusion of uncertainty, people must realize that they need to inhibit this automatic activation and reason in a more conscious, analytic manner (Evans 2002). Only by doing so can they recognize that with some conditional forms, it is not possible for any conclusion to be drawn with certainty; thus, these forms are indeterminate.

Models of conditional reasoning development

Three models of conditional reasoning have shaped our understanding of the development of this important skill; Moshman's (1990) model of metalogical development is one of them. According to this model, understanding of the logical basis for indeterminacy requires explicit awareness of one's own reasoning, or metalogic. Such awareness is only implicit in childhood but becomes increasingly explicit in the adolescent years; full understanding does not emerge before adolescence (Byrnes and Overton 1986; Markovits and Vachon 1990; Moshman and Franks 1986).

Another model, proposed by Henry Markovits (Markovits 1993, 2000; Markovits and Barrouillet 2002), emphasizes the role of semantic memory in affecting reasoning processes. According to Markovits (2000), people construct mental models of premises in part by activating relevant knowledge from long term memory (such as, in the example above, the role of lifting weights in building muscle strength). In the case of drawing uncertainty conclusions with indeterminate argument forms, reasoners must inhibit automatically activated inferences, and also access their semantic memories to generate alternatives to the antecedent in the major premise. With indeterminate forms, the activation of alternative antecedents ("If *a* then *q*,") to the premise, "If *p* then *q*," increases the probability that a reasoner will inhibit the invited inference and instead realize that no certain conclusion can be made. In the example above, if the reasoner can activate an alternative antecedent, such as "A person could also have strong muscles if she trains horses," the invited certain inference ("Harriet lifts weights regularly") is less likely to be made. Markovits' model (Markovits and Barrouillet 2002) also notes that when reasoning from premises that are inconsistent with their own knowledge, people must limit their access to long term memory to prevent their knowledge from interfering with reasoning based only on premise information.

A third approach, the dual-process theory of reasoning, has also been applied fruitfully to conditional reasoning problems. According to this approach, information processing operates in two systems, labeled by Klaczynski and Daniel (2005) as experiential (or heuristic) and analytic. The experiential system is largely automatic, relatively fast, and operates with minimal effort on meaningful tasks. In contrast, the analytic system is effortful, and operates under conscious control according to abstract rules of inference. It involves separating the logical structure of a task from its social contents, as well as inhibiting access to prior knowledge (Evans 2002; Klaczynski and Daniel 2005; Stanovich 1999; Stanovich and West 2000). The experiential system, because it is less effortful, is predominant in everyday problem-solving; analytic processing is more likely when motivation for a precise response is high or when a task clearly requires logical analysis.

Reasoning, reading, and metacognition

All three of the above approaches have made useful predictions about the development of logical reasoning, but they are also relevant to reading comprehension, particularly in adolescence and early adulthood. Secondary-level reading materials demand more inference and critical thinking, so they are more likely to elicit analytic processing. Conditional logical forms are modeled in children's reading materials with increasing frequency at higher grade levels (Franks et al. 1997), and are present in the hypotheses students read about in science texts. Comprehension of advanced reading materials may require readers not only to reason, but also to evaluate the reasoning modeled in texts.

That metacognition is fundamental to such evaluation is a point stressed by Baker (2004), who noted that the National Science Education Standards (NRC 1996, p. 42) include experiences with evaluation criteria because they help students apply standards of scientific practice to their own and others' scientific efforts. The internalization of such standards is viewed as critical to student achievement in science. In order to compare the conclusions presented in scientific texts with their own internalized standards, students must exercise metacognition.

Otero (2002) also stressed the importance of metacognition, and particularly comprehension regulation, to reading scientific material. He noted that in science education, readers often encounter information that cannot be comprehended because it is inconsistent with their knowledge; in such cases, adequate regulation is critical. Readers must be able to recognize comprehension difficulties and state them explicitly before repair inferences can be made. Otero also stressed awareness of uncertainty as crucial to scientific understanding. In the present study, both uncertainty and the relation of text information to readers' prior knowledge are explored in the context of logical reasoning about texts.

The Moshman (1990) model, the Markovits model (Markovits and Barrouillet 2002), and the dual process approach (Klaczynski and Daniel, 2005; Stanovich and West 2000), all emphasize processes that require awareness of one's own thinking, or metacognition. Moshman emphasizes metalogic, Markovits proposes the conscious accessing or blocking of access to semantic memory, and the dual process approach stresses the conscious control needed for analytic processing. Metacognition is also necessary for many reading tasks, particularly comprehension monitoring (Burton and Daneman 2007; Daneman and Hannon 2001; Kinnunen and Vauras 1995; Thiede et al. 2003). If readers neglect to monitor their own comprehension as they are reading, they will eventually experience comprehension failure as texts become more difficult. But according to Walczyk (1995) compensatory encoding model, many reading obstacles can be overcome metacognitively. If readers are unskilled at decoding, or if they have inefficient semantic memory access, they can slow their reading rate. If their verbal working memory skills are ineffective, they can look back in text more frequently. Looking back reinstates text information that has faded from working memory, but readers can only use this compensatory strategy if they are metacognitively aware of the need for it (Walczyk and Taylor 1996).

Looking back in text

Reading researchers have long been aware of the utility of looking back in text for improving reading comprehension. Early studies of this strategy (Alessi et al. 1979; Alverman 1988; Garner and Reis 1981; Zabrucky and Ratner 1986) found positive effects on comprehension for looking back as measured by readers' physical turning of pages. More recent studies (Burton and Daneman 2007; Kinnunen and Vauras 1995) have assessed looking back with eye tracking equipment that can record saccades (rapid forward eye movements averaging 6–9 character spaces) and regressions (rapid backward eye movements averaging 15–20 character spaces). Regressions, even for skilled readers, typically make up 10–15 % of saccades, and can result from either linguistic processing difficulties or oculomotor errors (Reichle et al. 2003). But since regressions involve backward movements of relatively few character spaces, a reader wishing to reinstate information from a previously read paragraph or page (e.g., in response to a question) would need to make several backward movements in the text. This kind of looking back is more conscious than regressions, and can be recorded by computerized tracking of readers' purposeful backward movements through texts presented on a screen (Walczyk and Taylor 1996), the method used in this study. Although looking back in text while reading may sometimes be automatic, Walczyk and Taylor (1996) argue that it is clearly metacognitive when readers have been instructed on the use of looking back as a compensatory mechanism. In the present study, readers were shown how to look back while reading on-screen texts, and were allowed to look back at will. Additionally, sentences were presented one at a time on the screen, so that in order to reinstate a previously read sentence, readers had to make a conscious decision to press a computer key and move backward in the text.

Rationale for the study

Understanding of the relations among reading, thinking, and cognitive development from early to late adolescence could yield important insights for improving students' reading skills in the middle and high school years. Consequently, the goal of this project was to explore cognitive processes involved in logical reasoning in a context that made them part of a reading task. Conditional logic premises were embedded in narratives that included plots, characters, and supporting details to facilitate readers' construction of mental models (Markovits and Barrouillet 2002). While reading the narratives aloud, readers periodically encountered questions in the texts that required logical deductions based on conditional premises; they were asked to answer these questions aloud as well (see Appendix 1 for excerpts from the narratives).

This format allowed us to observe the frequency with which readers looked back at previously read information before responding to logical deduction questions. We wanted to test the prediction that they would use this metacognitive behavior during reading comprehension tasks that required awareness of their own thinking, such as drawing indeterminate conclusions and reasoning from contrary to fact premises. Since metacognition is also needed for evaluation of texts, asking readers to evaluate a conclusion drawn by a story character should promote closer examination of the premises on which the character's conclusion is based, making lookbacks more likely. If so, lookback frequency should be greater among readers who evaluate characters' conclusions than among those who are only asked to generate their own conclusions.

In order to explore all these variables, we designed a reading task in which narrative texts that were factual (True Story), contrary to fact (False Story), or unverifiable via common world knowledge (Neutral Story) were embedded with conditional logic premises and deductive questions that required either determinate or indeterminate conclusions. Each reader was assigned to a different Response Task, and was asked either to draw logical conclusions (the Generate task) or to evaluate conclusions drawn by story characters that were either valid (the Valid task) or invalid (the Invalid task) according to standard logic.

Most studies of logical reasoning involve paper and pencil tasks, with analyses based on products (answers to questions). Although some studies also ask participants for justifications (e.g., Markovits 2000), the actual processes used in arriving at deductive conclusions are difficult to assess. Our participants' reasoning processes were observed in two ways. First, they were asked to "think aloud" while they read, and their comments, along with their responses to the logic questions, were recorded (comment analyses are reported elsewhere: Franks et al. 2011). Second, as reported here, metacognitive behavior was observed via the presentation of text to readers one sentence at a time on a computer screen, with readers controlling their own movements forward or backward through the text. A computer program (described below) recorded the number of times readers looked back in text in response to each question. Thus, we analyzed performance with two dependent variables: lookback frequency and logical conclusions.

Predictions of models: lookback frequency

Predictions based on the three models of logical reasoning development described above (Klaczynski and Daniel, 2005; Markovits and Barrouillet 2002; Moshman 1990) were assessed. According to Moshman (1990) explicit awareness of one's own reasoning increases with age. Increased amounts of looking back in text would suggest more explicit metacognitive awareness; the model would thus predict more looking back in text with increasing age, as well as more looking back on reasoning tasks requiring metalogic.

Markovits' model (Markovits and Barrouillet 2002) suggests that reasoning from contrary to fact premises requires that people limit access to long term memory, because prior knowledge can interfere with reasoning (Markovits and Vachon 1989; Markovits 1995). In our reading task, the differing cognitive demands presented by True, False, and Neutral premise information suggest that lookback frequencies could be differentially affected by story content.

As have other studies based on a dual-process approach (Klaczynski et al. 2004; Klaczynski and Daniel 2005; Daniel and Klaczynski 2006), we explored performance with both determinate and indeterminate conditional forms; determinate forms are thought to involve mainly experiential or heuristic processing, while indeterminate forms appear to require analytic processing. If this is so, then indeterminate forms should elicit more looking back in text, presumably to check the exact wording of premises.

Predictions of models: logical conclusions

With regard to the logical conclusions drawn by readers, all three of the models discussed above, and many studies of conditional reasoning (Byrnes and Overton 1986; Chao and Cheng 2000; Daniel and Klaczynski 2006; Franks 1996, 1997; Klaczynski 2001; Markovits and Vachon 1990) support an expectation of increases with age in arriving at conclusions that are consistent with standard logic ("yes" for MP, "no" for MT, and "uncertain" for AC and DA). Because AC and DA (hereinafter referred to as Indeterminate forms) require the more effortful analytic processing, they remain more challenging than MP and MT (hereinafter referred to as Determinate forms), even for adults, and valid conclusions with Indeterminate forms can be expected to be less frequent than with Determinate forms.

Moshman (1990) proposed that awareness of one's own reasoning is only implicit in childhood and becomes more explicit in adolescence. If so, the modeling of correct conclusions by story characters in text may facilitate readers' conclusions on tasks requiring metalogic. They may, for example, find it easier to agree with uncertainty conclusions on AC and DA when story characters model such conclusions (a task requiring only implicit metalogical awareness) than to generate uncertainty conclusions themselves or to recognize when story characters' modeled certainty conclusions on AC and DA are not valid (tasks requiring more explicit metalogical awareness).

Finally, many logical reasoning studies (Franks 1996, 1997; Markovits and Vachon 1989, 1990; Markovits 1995; Markovits et al. 1996; Moshman and Franks 1986), particularly those based on Markovits' model, have found logical conclusions to be affected by reasoners' prior knowledge. Performance is usually better with premises regarded as true than with premises regarded as false, because in the latter case, reasoners must limit access to knowledge stored in long-term memory and focus only on the relations specified by the premises.

In summary, the three models of conditional reasoning discussed here make related predictions regarding metacognitive behavior. Based on these models, looking back in text can be seen as an indicator of both metacognitive awareness and the analytic processing such awareness allows. Our purpose in exploring all three models was not to establish support for any one model over the others. Instead, we used their complementary predictions to explore how readers *reason* about the texts they read.

Research hypotheses

Based on the three theoretical perspectives described above, we made several predictions for the two dependent variables explored in this study, lookback frequencies and logical conclusions. Predictions are numbered for reference in the analyses.

Predictions for lookback frequencies

- P1. Consistent with Moshman's proposal that metacognitive awareness increases with age (Moshman 1990), we predicted a general increase in lookback frequency with age. We also predicted that age would interact with Response Task, such that lookback frequencies would be more affected by Response Task among older students than among younger ones, resulting in more lookbacks for the evaluation tasks among older students.
- P2. We predicted that overall lookback frequency would be affected by Response Task, with the two evaluation conditions eliciting more looking back than the generate condition. We further predicted that evaluating invalid logical conclusions would elicit more looking back than evaluating valid conclusions. These predictions are consistent with both Moshman's (1990) model and the dual processing approach, since evaluation of reasoning in texts involves both metalogic and analytic processing.
- P3. Consistent with the dual-process theory view that indeterminate forms require analytic processing (Klasczynski and Daniel, 1995) we predicted that lookback frequencies would be greater with Indeterminate forms (AC, DA) than with Determinate forms (MP, MT).
- P4. Consistent with the Markovits model (Markovits and Barrouillet 2002), we predicted that lookback frequencies would be affected by story content (True, False, Neutral) and that lookbacks would be greatest on the False story, where readers were required to block access to their long-term memories and reason solely on the basis of premise information.

Predictions for logical conclusions

P5. Consistent with all three models, as noted earlier, we predicted that readers' valid conclusions (i.e., those consistent with standard logic) would increase with age.

P6. We predicted that overall performance would be highest on the Valid task, where valid logical conclusions were modeled, and lowest on the Invalid task, where invalid logical conclusions were modeled. We also predicted that with Indeterminate forms, students would be more likely to support a valid uncertainty conclusion modeled in text (Valid task) than to arrive at such a conclusion independently (Generate task) or to recognize an invalid certainty conclusion (Invalid task).

P7. We predicted that students would arrive at more correct conclusions with Determinate forms than with Indeterminate forms. This prediction is consistent with all three models and with the research literature in conditional reasoning.

P8. We predicted that students' valid conclusions would be influenced by story content (True, False, or Neutral) on all three Response Tasks, and that performance with the False content would be lower than performance with True and Neutral content.

Method

Participants

Participants were students in the 5th grade (N=22; Mean age 11.2 years), 8th grade (N=30; Mean age 13.8), 10th grade (N=21; Mean age 15.7), and College (N=23; Mean age 21.5). Students in the 5th, 8th, and 10th grades were all average to above-average readers by teacher nomination. All were students in a University-affiliated laboratory school whose student population is demographically similar to the local population. College students were drawn from a participation pool of students in undergraduate Educational Psychology classes at a large Southeastern U. S. university, and participated as a course requirement.

Materials

Three narrative passages, 110 to 112 sentences in length, were used in the reading task. Each passage contained premise information in the four conditional logic forms, with two Determinate forms (MP and MT) and two Indeterminate forms (AC and DA), followed by a question requiring a deductive conclusion. As noted earlier, when reasoning about Indeterminate forms, people are more likely to inhibit "invited" certain conclusions if they are able to activate alternative antecedents ("If *a* then *q*,") to the premise, "If *p* then *q*." This activation allows them to realize that no certain conclusion can be drawn, since "*a*" and "*p*" are equally likely. To facilitate such a realization in the reading context, Indeterminate forms in each story were accompanied in the text by alternate antecedents, that is, alternates to "*p*" that were expressed as equally likely possibilities. (See Appendix 1 for examples in excerpts from the three stories. In each story, the major and minor premises of each logical form are underlined, and the alternate premises for Indeterminate forms are in italics. The first underlined sentence in each excerpt presents the first premise of the logical argument; that is, it presents the "if *p* then *q*" claim (e.g., "If it's a garter snake, it will have stripes."). In each story, this sentence is followed by two filler sentences. The next sentence presents the second premise of the logical argument, that is, it instantiates *q* or *p* (e.g., "That was not a garter snake."). These markings were not present in the computer-screen texts). In all cases, the distance between the first and second premises of the logical argument was kept constant.

Premise information in the three stories was factual (True Story), contrary to fact (False Story), or unverifiable (Neutral Story) via common world knowledge (see Appendix 1). In the True story, which described a school field trip to a nature center, readers could access their own knowledge in drawing conclusions. In the Neutral story, which described a mission on a fictitious spaceship, premise information was unrelated to common world knowledge. In the False story, which described a fantasy town in the desert, premise information contradicted common world knowledge, so readers needed to inhibit access to their long-term memories in order to draw conclusions. Stories were presented in staggered order in each condition at each age level to avoid effects of order of presentation. For each form, the premise information was followed by a question that readers answered aloud. Each story also included two "filler" questions that required simple, non-deductive conclusions. All three stories were composed at a 5th grade reading level, based on the Flesch-Kincaid Grade Level formula (Flesch and Lass 1996).

Procedures

All participants were asked to read and "think aloud" on three narrative passages, presented one sentence at a time on a computer screen. Students were audio taped as they read, and they controlled the presentation of sentences with forward and backward arrow keys. Each test session lasted approximately one hour.

Each participant was assigned to one of three different Response Tasks. On the Generate task, readers were simply asked to generate logical conclusions based on premises in the stories, as they are in most studies of logical reasoning. On the Valid task, the narratives included conclusions drawn by story characters, all of which were logically valid according to standard logic, and readers were asked to evaluate the characters' conclusions. On the Invalid task, the same story characters drew logically invalid conclusions to the same questions, and readers were asked to evaluate those conclusions.

All elements of the experiment were controlled using E-prime software (Schneider et al. 2002). The text and questions were presented, left justified, on a computer monitor using 24 point Times New Roman font. Participants were instructed to use the right arrow (forward) key to read the next sentence and the left arrow (backward) key to go back in the text. The number of key presses and the time stamp of those presses were used to confirm how frequently readers reinstated information in working memory before answering each question. In all three conditions, participants were first asked to read a brief "practice" story and to answer questions of the same type (either generating or evaluating conclusions) as the questions in the test stories. The practice story ensured that all students understood how to move forward and backward in the texts as necessary, and included a question for which no certain conclusion could be drawn, to make it clear that uncertainty was an acceptable response.

Assessment of dependent variables

The first variable assessed was lookback frequency; the total number of lookbacks in response to each question was recorded for each participant, using the software described above. Lookbacks were recorded for four logic questions and two filler questions in each of three stories, but only the results for the logic questions were analyzed. The second variable assessed was the number of correct responses to the questions given by each participant.

Verbal responses to each question were recorded with digital audio recorders and later transcribed. Scores analyzed were number of correct responses to four logic questions in each of three stories, for a total of 12 possible, and total number of correct responses on Determinate and Indeterminate forms in all three stories, with 2 Determinate and 2 Indeterminate questions in each story, for a total of 6 possible for each form on all three stories.

For each dependent variable, we performed two analyses: 1) a 4x3x3 mixed factors ANOVA, with Age (5th grade, 8th grade, 10th grade, and College) and Response Task (generate conclusion, evaluate valid conclusion, and evaluate invalid conclusion) as between-subjects factors and Story Content (False, True, Neutral) as a within-subjects factor, and 2) a 4x3x2 mixed factors ANOVA, with Age (5th grade, 8th grade, 10th grade, and College) and Response Task (generate conclusion, evaluate valid conclusion, and evaluate invalid conclusion) as between-subjects factors and Form (Determinate, Indeterminate) as a within-subjects factor. Thus, we considered the effects of Response Task on readers' metacognitive behavior (looking back) and on their reasoning with contrary to fact premises, factual premises, and neutral premises. We also explored the effects of Response Task on readers' looking back and on their reasoning with argument forms that require certain conclusions, and forms that require uncertain conclusions.

Results

Analyses of lookback frequencies

Effects of grade on lookback frequency

Table 1 (see Appendix 2) illustrates the mean lookback frequencies for each grade, response task, and story type in our sample. Contrary to our first prediction (P1), when means for total number of lookbacks at all four grade levels were compared, no significant main effect for grade was observed. Students in 5th grade, 8th grade, 10th grade, and College were not significantly different in their overall lookback frequency patterns.

We also predicted that age would interact with Response Task (P1), expecting that, due to their greater metacognitive awareness, children in 8th and 10th grade, as well as College students, would be more likely than 5th graders to look back on the evaluation tasks (Valid and Invalid) than on the Generate task. As illustrated by Fig.1, a significant interaction between grade and Response Task was observed in the False story only, $F(6, 84)=3.05$, $p=.009$; $\eta_p^2=.179$. The interaction was not as predicted, however; followup analyses indicated a significant effect for Response Task on the False story for the 5th graders, $F(2, 19)=6.31$, $p=.008$), the 8th graders, $F(2, 27)=16.25$, $p=.001$), and the 10th graders, $F(2, 18)=7.65$, $p=.004$), but not for the College students. Pairwise comparisons for the three Response Tasks in each grade indicated a significant difference in lookbacks only between the Generate task and the Valid task for the 5th graders. For 8th and 10th graders, differences were significant between Generate and Invalid, and also between Valid and Invalid, but not between Generate and Valid. The Bonferroni correction for a large number of comparisons was used for all pairwise comparisons in all analyses.

College students' lookback frequencies were not significantly different between the two evaluation tasks (Valid and Invalid) but indicated a greater general tendency to look back on all three Response Tasks (see Fig. 1, and Table 1 in Appendix 2 for means and standard deviations). This interaction, while not as predicted, represents a

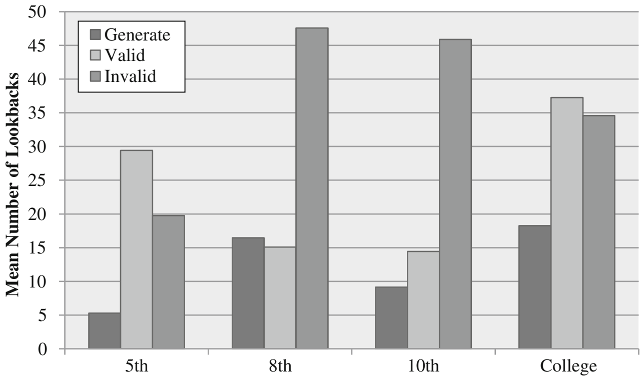


Fig. 1 Mean lookback frequencies on the False story for each grade on each of three response tasks

pattern found in several other analyses, reported below, wherein Response Task affected lookback frequency for younger readers, but not college students.

Effects of response task on lookback frequency

Consistent with our second prediction (P2) and as illustrated in Fig. 2, Response Task had a significant effect on the total number of lookbacks made by students, $F(2, 84)=11.67$, $p=.001$; $\eta_p^2=.217$.

Response Task also affected lookback frequency significantly with all three story contents (False story, $F(2, 84)=14.65$, $p=.001$; True story, $F(2, 84)=6.47$, $p=.002$; $\eta_p^2=.133$); Neutral story, $F(2)=7.88$, $p=.001$; $\eta_p^2=.158$), but interacted with age on the False story, as described above. Also as predicted (P2), for total number of lookbacks, and for all three stories, students exhibited the greatest amount of looking back when they were asked to evaluate logically invalid conclusions drawn by story characters (Invalid task) and the least amount when generating conclusions on their own (Generate task); see Fig. 2. Pairwise comparisons indicated that lookback

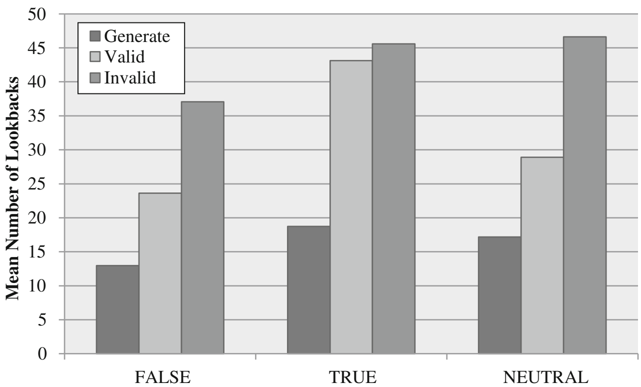


Fig. 2 Mean lookback frequencies for the total sample on each of three response tasks for the False Story, the True Story, and the Neutral Story

frequencies were significantly greater on the Invalid task than on the Valid task for all three stories and for total lookbacks, and were also significantly greater on the Valid task than on the Generate task for the False and True stories and for total lookbacks.

Between the two evaluation tasks, lookback frequencies were significantly greater for the Invalid task than for the Valid task for the False and Neutral stories, but not for the True story or for total lookbacks (See Table 1 in Appendix 2 for means and standard deviations). As we predicted (P2), lookback frequencies were always highest for the Invalid task, though the difference between the Valid and Invalid tasks was not always significant. Also as predicted (with the exception of the Valid task with the Neutral story), evaluating conclusions modeled by story characters in text elicited more looking back by readers than did generating their own conclusions (P2).

Effects of form (Determinate vs. Indeterminate) on lookback frequency

We predicted a main effect for Form (P3), expecting lookback frequencies to be greater with Indeterminate forms (AC, DA) than with Determinate forms (MP, MT), due to the analytic processing required by the former. Our analysis revealed two significant interactions, however, one between Form and Response Task, $F(2, 84)=3.16$, $p=.047$; $\eta_p^2=.070$, and one between Response Task and Grade, $F(6, 84)=2.51$, $p=.028$; $\eta_p^2=.152$. Followup analyses of the Form x Response Task interaction revealed significant main effects for Response Task with both Determinate forms, $F(2, 84)=8.01$, $p=.001$, and Indeterminate forms, $F(2, 84)=13.90$, $p=.001$, but different relationships among the three reading conditions were found with the two different kinds of argument forms (see Fig. 3). With Determinate forms, lookback frequency was significantly greater in both of the evaluation tasks (Valid and Invalid) than on the Generate task, but means for the two evaluation tasks were not significantly different. In contrast, with Indeterminate forms, lookback frequency was significantly greater on the Invalid task than on either the Generate task or the Valid tasks, while those two tasks did not produce significantly different lookback frequencies. Our prediction (P2) that readers would look back significantly more when evaluating invalid conclusions than when evaluating valid conclusions was supported only when the conclusions in question were based on Indeterminate forms, which require an uncertainty conclusion. With Determinate forms (requiring a certainty conclusion), the sample as a whole did not show a significant difference in looking back when evaluating either valid or invalid conclusions (but see Response Task x Grade interaction below).

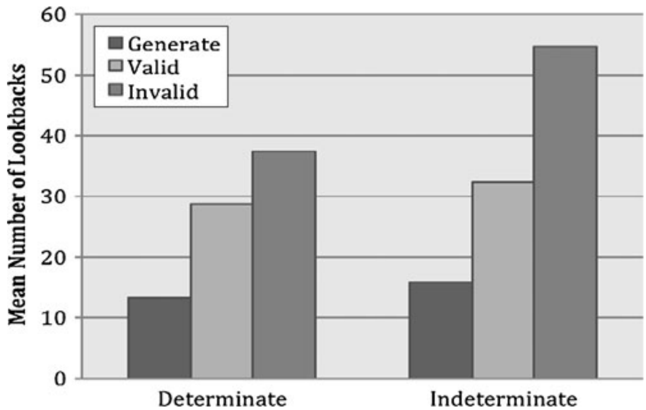


Fig. 3 Mean lookback frequencies with Determinate and Indeterminate forms on each of three response tasks

These results suggest that with Determinate forms, evaluating conclusions elicited more looking back than generating conclusions, regardless of whether readers were asked to evaluate logically valid or logically invalid conclusions. With the more difficult Indeterminate forms, evaluating logically invalid conclusions prompted significantly more looking back than evaluating logically valid conclusions.

The results above reflect findings for the whole sample, collapsed across age. But a significant Response Task by Grade interaction was also found in this analysis, $F(6, 84)=2.51$, $p=.028$; $\eta_p^2=.152$. The effects of Response Task and Grade were therefore explored separately for performance with the Determinate and the Indeterminate forms. With Indeterminate forms, there was a significant effect for grade, $F(3, 84)=2.73$, $p=.049$, but no significant Grade by Response Task interaction. Pairwise comparisons indicated a significant difference in total number of lookbacks only between the 5th graders (Mean=21.22) and the College students (Mean=44.21).

With the Determinate forms, however, a significant Response Task by Grade interaction was observed, $F(6, 84)=2.42$, $p=.033$, such that significant effects of Response Task with Determinate forms were present in the 5th grade, $F(2, 19)=4.03$, $p=.035$, the 8th grade, $F(2, 27)=4.36$, $p=.023$, and the 10th grade, $F(2, 18)=5.25$, $p=.016$, but not among the College students. Pairwise comparisons with the Determinate forms indicated that fifth graders looked back significantly more on the Valid task than on the Generate task, while eighth graders looked back significantly more on the Invalid task than on the Generate task. Tenth graders looked back significantly more on the Invalid task than on either the Valid task or the Generate task (see Fig. 4). We predicted that readers would show significantly higher levels of metacognitive behavior when evaluating logically invalid inferences (Invalid task) than on the other Response Tasks (P2). As Fig. 4 illustrates, this tendency was not present among 5th graders, was strong among 8th and 10th graders, but had diminished by the college years. Our prediction (P2) was supported with Indeterminate forms for the total sample, and supported with Determinate forms for students in middle adolescence, but not for 5th graders or college students.

Effects of story content on lookback frequencies

We predicted that lookback frequencies would be affected by story content, and that the False story would elicit the most looking back, since it required readers to ignore their prior

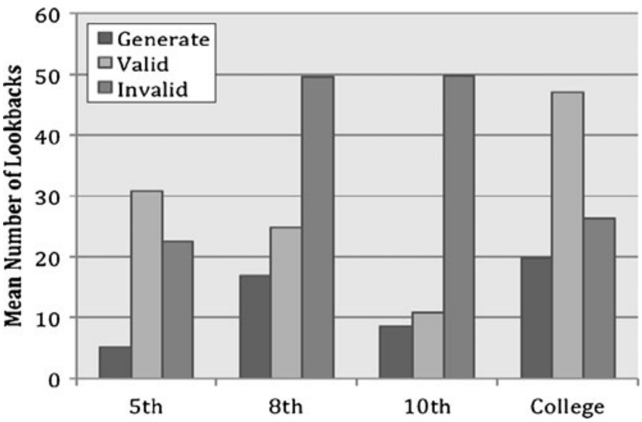


Fig. 4 Mean lookback frequencies with Determinate forms for each grade on each of three response tasks

knowledge and focus only on the premises (P4). This prediction was not supported, but a significant three-way interaction of Story Content, Grade, and Response Task, $F(12, 84)=2.14$, $p=.017$; $\eta_p^2=.132$) was observed. While not predicted, this interaction revealed an interesting developmental trend.

Followup analyses of the three-way interaction illustrated the different ways in which Story Content and Response Task interacted at the different grade levels. For the 5th graders, the effect of Response Task was significant on all three stories (False story, $F(2, 19)=6.31$, $p=.008$; True story, $F(2, 19)=4.52$, $p=.025$; Neutral story, $F(2, 19)=3.73$, $p=.043$). Pairwise comparisons indicated significant differences between the Generate task and the Valid task on the False story, between the Generate task and the Invalid task on the True story, and between the Generate task and the Invalid task on the Neutral story, with the lowest means on the Generate task (see Table 1, Appendix 2).

For the 8th graders, the effect of Response Task was significant on two stories (False story, $F(2, 27)=16.2$, $p=.001$; Neutral story, $F(2, 27)=11.05$, $p=.001$). Significant differences were observed between the Generate task and the Invalid task, and between the Valid and the Invalid tasks, on both the False and the Neutral stories, with the lowest means on the Generate task (see Table 1, Appendix 2).

For the 10th graders, the effect of Response Task was significant on only one story, the False story, $F(2, 18)=7.64$, $p=.004$, with significant differences between the Generate task and the Valid task, and between the Valid and the Invalid tasks, with the lowest means on the Generate task and the highest on the Invalid task (see Table 1, Appendix 2).

College students' lookbacks were least affected by Response Task, with no significant differences among the tasks on any story. Taken together, these results suggest that the effect of Response Task on the tendency to look back in text, while strong among 5th graders, diminished among older adolescents and was not observable by young adulthood. This gradual decline in the effect of Response Task, accompanied by a greater general tendency to look back in text in College students, suggests that College students may use looking back as a general metacognitive strategy, regardless of task demands. It is also interesting to note that the Response Task effect diminished earliest (after 5th grade) for the True story, in which readers could access their own stored knowledge to help them make inferences, and remained longest (through the 10th grade) for the False story, which required readers to limit their access to stored knowledge. This is consistent with Markovits' model (Markovits and Barrouillet 2002), even though the actual rate of looking back was not highest in the False story.

Summary of lookback frequency results

The results largely supported the prediction (P2) that Response Task would affect overall lookback frequency, with the two evaluation tasks eliciting more looking back than the generation task. The prediction (P2) that evaluating invalid logical conclusions would elicit more looking back than evaluating valid conclusions was supported with Indeterminate forms for the whole sample, and was also supported with Determinate forms among students in middle adolescence (8th and 10th graders). Finally, the prediction (P3) that Indeterminate forms, which require analytic processing, would elicit more looking back than Determinate forms, was supported only when students were evaluating invalid conclusions (Invalid task). With regard to Story Content, the prediction (P4) that the False story would elicit the most looking back was not supported. Followup to the significant three-way interaction of Story Content, Grade, and Response Taks, however, suggested a developmental trend wherein the effect of Response Task on lookback frequency was strong among 5th graders, diminished among older adolescents, and was not observable by young adulthood. College students' lookback frequencies were not affected by Response Task, but were greater overall than those of younger students, suggesting that College students use looking back as a general metacognitive strategy.

Analysis of logical conclusions

Several researchers (Evans 2002; Daniel and Klaczynski 2006) have criticized the practice of judging conclusions as "correct" or "incorrect," since people's everyday conditional reasoning may be inconsistent with standard logic and nevertheless correct in terms of natural language interpretations of the conditional. For the purpose of this study, reference to conclusions as "correct" means only that they are consistent with standard logic. For each Response Task in this study, a "correct" answer required a different response. On the Generate task, students needed to generate responses that were valid, that is, consistent with standard logic (e.g., "yes" for MP, "no" for MT, and "uncertain" for AC and DA). On the Valid task, a "correct" response meant agreeing with conclusions modeled in the text, all of which were valid, or consistent with standard logic, while on the Invalid task a "correct" response meant disagreeing with conclusions modeled in the text, all of which were invalid, or inconsistent with standard logic (e.g., "no" for MP, "yes" for MT, "yes" for AC, and "no" for DA). Students' responses to the four logic questions in each of the three stories were analyzed for total scores within each story (possible 4 correct) and also for their total scores on all three stories (possible 12 correct). Table 3 (see Appendix 2) illustrates their correct responses by grade, response task, and story content.

Effects of grade on logical conclusions

Because the three Response Tasks required different "correct" responses, effects on logical conclusions were explored separately for each Response Task, beginning with grade effects. No main effects for grade were observed on either the Valid task or the Invalid task, nor did grade interact with any other variables on those tasks. On the Generate task, however, a significant main effect for grade was observed, $F(3, 27)=6.25$, $p=.002$, with $\eta_p^2=.409$. Pairwise comparisons showed College students scored significantly higher than students at all other grade levels (see Table 3, Appendix 2). Our prediction (P5) that correct conclusions would increase with age was not broadly supported; only on the Generate task were any age effects found.

Effects of response task on logical conclusions

As predicted (P6), significant main effects for Response Task were found with all three story contents (False story, $F(2, 76)=9.09$, $p=.001$, $\eta_p^2=.239$; True story, $F(2, 76)=11.72$, $p=.001$), $\eta_p^2=.235$; Neutral story, $F(2, 76)=26.49$, $p=.001$), $\eta_p^2=.410$, as well as for total scores, $F(2, 76)=31.11$, $p=.001$, $\eta_p^2=.450$. In all cases, pairwise comparisons revealed the same pattern: scores on the Valid task, where students evaluated logically valid conclusions in text, were significantly higher than scores on both the Generate task (where students generated their own conclusions) and on the Invalid task (where students evaluated logically invalid conclusions in text) (see Table 3 in Appendix). As noted below, however, this effect of Response Task on readers' valid conclusions was entirely the result of their performance with Indeterminate forms.

Effects of form (determinate vs. Indeterminate) on logical conclusions

Students' responses to questions on the Determinate forms (MT and MP) were compared with their responses to questions on the Indeterminate (DA and AC) forms (collapsed across the three stories). We predicted that performance with Determinate forms would be higher than performance with Indeterminate forms across all grades and on all Response Tasks (P7). Instead, a significant interaction between Form and Response Task, $F(2, 76)=26.21$, $p=.001$; $\eta_p^2=.407$, indicated different effects of Response Task with the Determinate and Indeterminate forms (see Fig. 5, and Table 4 in Appendix 2).

As Fig. 5 illustrates, followup analyses to the interaction indicated that students' logical conclusions with Determinate forms were similar on all three Response Tasks. There were no significant main effects or interactions for Response Task or Grade. With Indeterminate forms, however, there were significant main effects for Response Task, $F(2, 76)=46.53$, $p<.001$, and Grade, $F(3, 76)=2.64$, $p=.050$, but no interaction of Response Task and Grade. Pairwise comparisons for the Grade effect indicated College students scored significantly better than students at all other grade levels with Indeterminate forms (see Table 4).

Pairwise comparisons for Response Task showed that with Indeterminate forms, scores on the Valid task were significantly higher than scores on both the Generate and the Invalid tasks, which were not significantly different (see Fig. 5, and Table 4 in Appendix 2). As Fig. 5 also illustrates, scores were significantly higher with Determinate than with Indeterminate forms on the Generate task and on the Invalid task, but not significantly different on the Valid task. Readers as a whole found it almost equally difficult either to generate logically valid uncertainty conclusions themselves (Generate task) or to reject the invited but logically invalid certainty conclusions modeled on the Invalid task. But on the Valid task, they were able to accept as correct the logically valid uncertainty conclusions to the same questions, when they were modeled by story characters.

Effects of story content on logical conclusions

We predicted (P8) that Story Content would affect conclusions on all three Response Tasks, and that scores in the False story would be the lowest. Instead, a significant interaction of

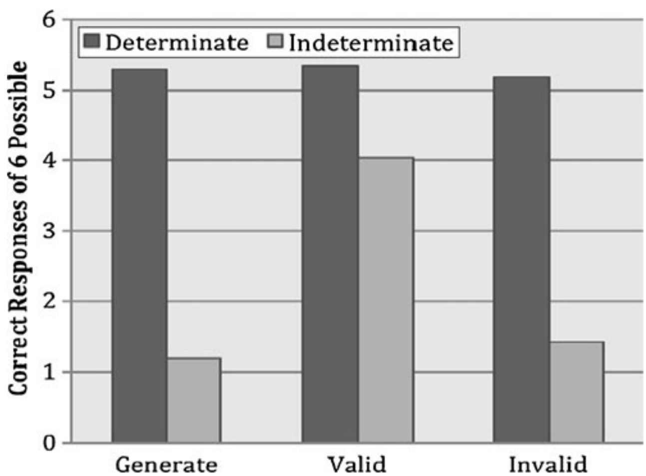


Fig. 5 Mean number of correct responses to logic questions with Determinate and Indeterminate forms on each of three response tasks

Story Content and Response Task was observed, $F(4, 76)=3.87$, $p=.005$; $\eta_p^2=.092$), and Story effects were explored separately for each Response Task. On both the Valid and the Invalid tasks, the effect for Story Content was significant (Valid task, $F(2, 25)=6.57$, $p=.003$; Invalid task, $F(2, 24)=14.01$, $p=.001$). As illustrated by Fig. 6, on the Valid task, scores on both the True and Neutral stories were significantly higher than scores on the False story. On the Invalid task, scores on the True story were significantly higher than scores on both the False and the Neutral stories. Our prediction (P8) was not supported on the Generate task, where no significant effect for Story Content was observed. Grade did not interact with Story Content on any Response Task.

Students' stronger performance with True content on the Invalid task suggests that it was easier for readers to reject a modeled (but invalid) certainty response to AC or DA when they were reasoning from texts that allowed them to draw on their prior knowledge, something they could not do with either False or Neutral content. The effects for Story Content on both the Valid and the Invalid tasks show that evaluating story characters' conclusions is particularly difficult when, as with the False story, readers must limit access to their own prior knowledge and reason from empirically false premises.

Summary of logical conclusions results

The prediction (P5) that correct conclusions would increase with age was not strongly supported; age effects were found only on the Generate task and only indicated better performance by College students than all other grades. The prediction (P6) that scores on the Valid task would be higher than on the other two tasks was supported, but only with Indeterminate forms; with Determinate forms, students performed equally well on all three Response Tasks. Consistent with our prediction (P7), performance with Indeterminate forms (requiring an uncertainty conclusion) was consistently and significantly lower than performance with Determinate forms (requiring a certainty conclusion). As noted above, the prediction (P8) that Story Content would affect logical conclusions was not supported on the generation task, but was

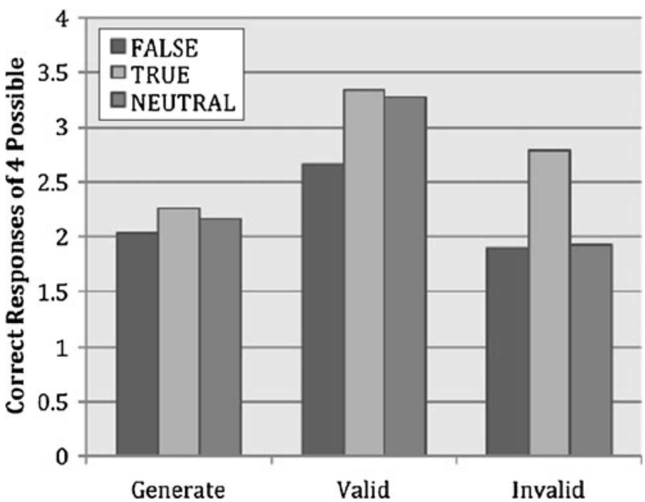


Fig. 6 Mean number of correct responses to logic questions in False, True, and Neutral stories on each of three response tasks

supported on the two evaluation tasks; performance with False content was generally lower, and never significantly higher, than performance with True or Neutral content.

Discussion

Our results inform the study of the comprehension of complex texts, because they demonstrate the relevance of models of logical reasoning to the metacognitive behaviors involved in reasoning about such texts. The study also offers a contribution to the logical reasoning literature, because it examines the way people *process* conditional reasoning problems as well as their responses to them. In that they are often consistent with more than one model, our findings illustrate the complementary nature of these models in characterizing logical reasoning.

Lookback frequency

The effects of Response Task on looking back have implications for both the dual-process model and Moshman's model. We expected that the metacognitive behavior of looking back in text would increase under conditions requiring the most analytic processing (dual process model) or metalogic (Moshman 1990). As predicted by both models (P2), being asked to evaluate, rather than generate logical conclusions elicited a much greater tendency on readers' part to look back at previously read text before answering.

The developmental trend suggested by the interactions of age with other variables offers further implications for Moshman's model. The Response Task effects with Determinate forms and on the False story (for students in grades 5, 8, and 10, but not for College students) as well as the three-way interaction of Grade, Response Task, and Story Content all point to a common effect: Response Task affected lookback frequency significantly for the three youngest groups of readers, but not for the College students. Taken together, these findings suggest that in early and middle adolescence, students exhibited different frequencies of metacognitive behavior depending on whether they were asked to generate their own conclusions or to evaluate conclusions modeled in text. College students were not so affected by Response Task, but looked back more overall than the other age groups. These results offer modest support to Moshman's (1990) proposal that metacognitive awareness increases with age. Students appear to have become proficient at this metacognitive behavior by the time they are in college, and to apply it more generally as a useful comprehension strategy, rather than differentially with different task demands.

Again based on both the dual-process approach and Moshman's (1990) model, we also expected (P2) that evaluating invalid conclusions would elicit more looking back than evaluating valid conclusions. With the Determinate forms, this effect was demonstrated only for 8th and 10th grade students. With the more difficult Indeterminate forms, all readers looked back significantly more when evaluating logically invalid than logically valid conclusions. This is important because the logically invalid Indeterminate inferences modeled in the text all expressed the most common "invited" inferences with these forms (a certain "yes" conclusion for AC and a certain "no" conclusion for DA). In order to reject the conclusions modeled in the texts, readers had to reject these invited inferences, even though story characters explicitly modeled them. This would require analytical processing, and likely include increased attention to the wording of premise information (Klaczynski 2000).

The even greater frequency with which participants looked back in text when evaluating logically invalid Indeterminate inferences suggests that readers gave careful scrutiny to these invited inferences before deciding whether or not the story characters who made them had made correct decisions. This finding is particularly relevant to the dual-process approach because it offers behavioral evidence that readers process logical forms requiring an uncertainty conclusion differently than forms requiring a certainty conclusion, at least when those forms are embedded in a full text. Even when students failed to reject the invalid conclusions, they clearly did not do so without consideration. Their responses, therefore, were unlikely to be based on the quick, automatically activated inferences used in heuristic processing.

Finally, the effects of Story Content on looking back have minor implications at best for the Markovits model. According to this model (Simoneau and Markovits 2003), drawing conclusions based on false premises requires readers to block access to their long-term memories, because their stored world knowledge contradicts the false premises and thus interferes with deductive conclusions. As such, the False story in this study arguably required the most metacognitive awareness, and we predicted greater lookback frequencies with this story (P4). But while 8th and 10th graders exhibited the expected behavior (see Fig. 2), 5th graders and College students did not, so this prediction was not strongly supported.

Logical conclusions

As we predicted (P6), scores on the Valid task, where readers evaluated logically valid conclusions drawn by story characters, were significantly higher than scores on the other two Response Tasks. Since this effect was entirely the result of performance with Indeterminate forms, however, it is clear that the effects of generating vs. evaluating conclusions are strongly influenced by the difficulty of the argument form readers are reasoning about. The effects of logical form itself (Determinate vs. Indeterminate) on logical conclusions illustrate this further. Students' performance on the logic questions indicated that, as predicted by all three models (P7), they found Indeterminate forms more difficult than Determinate forms. Also as predicted (P6), despite high rates of looking back on the Invalid task, when asked to evaluate story characters' (logically invalid) certainty conclusions on AC and DA, readers of all ages were unlikely to disagree with the story characters and draw uncertainty conclusions (averaging fewer than 2 out of 6 possible correct responses). This was true even though the texts provided alternative antecedents, which other research has found to increase the probability of uncertainty conclusions (Daniel and Klaczynski 2006; Janveau-Brennan and Markovits 1999). Readers were equally unlikely to arrive at uncertainty conclusions on their own on the Generate task. These results are consistent with the dual processing model, which predicts greater difficulty with Indeterminate forms, but they have particular relevance for Moshman's (1990) model when compared with the results observed on the Valid task.

In contrast to readers assigned to the Invalid task, readers assigned to the Valid task were much more likely to agree with story characters' (logically valid) uncertain conclusions on AC and DA (averaging 4 out of 6 possible correct responses). Modeled uncertainty conclusions in text seemed to be accepted in this context, even though readers asked to generate such conclusions on their own, or to reject invalid certainty conclusions modeled in text, were largely unable to do so. Rather (and consistent with Moshman's (1990) model), readers assigned to the Valid task showed *implicit* understanding in agreeing with uncertainty conclusions when story characters modeled them. Moshman's (1990) model also proposes that such understanding becomes more explicit with age, an effect not found in our results.

The lack of strong age effects on logical conclusions was the most unexpected finding in this study. Our prediction (P5) that correct conclusions would increase with age was not strongly supported; only in Condition 1 were any age effects found. This condition most closely resembles other studies of conditional reasoning in the sense that students were asked to generate conclusions rather than evaluate them; however, other studies (Byrnes and Overton 1986; Chao and Cheng 2000; Daniel and Klaczynski 2006; Klaczynski 2001; Markovits and Vachon 1990) have typically found significant differences in performance between younger and older children. In contrast, we found age differences only between the College students and all other readers. While these findings are not consistent with predictions made by any of the three models described here, they may simply be the result of differences between our reading task and experimental conditions and those used in earlier work based on these models.

Finally, our results with the effect of Story Content on logical conclusions were somewhat consistent with Markovits' model, but not overwhelmingly so. Our prediction (P8) that Story Content would affect logical conclusions was not supported on the Generate task (see Fig. 6). This was unexpected, because the Generate task most resembles the bulk of studies in conditional reasoning, where participants are only asked to draw conclusions, not evaluate them. In studies of this sort (Franks 1996, 1997; Markovits and Vachon 1989, 1990; Markovits 1995; Markovits et al. 1996; Moshman and Franks 1986) reasoning from False premises typically results in significantly lower performance.

The predicted Story Content effects were observed in the evaluation conditions, however, and performance with False content, while not always significantly lower, was never significantly higher than performance with True or Neutral content. The effects for Story Type on both the Valid and the Invalid task show that evaluating story characters' conclusions is more difficult when readers must limit access to their own prior knowledge in order to reason from empirically false premises. The stronger performance with True content also suggests that it is easier for readers to reject a modeled (but invalid) certainty response to AC or DA when they can draw on their prior knowledge in the process.

Limitations of the study

Because of the large amount of time needed to collect data in this experiment (students were also "thinking aloud" as they read the stories and answered questions), the number of participants in each condition was relatively small. All these results, therefore, must be interpreted with caution. Lookback frequency was a dependent variable in this study, since our purpose was to observe the effects of different response tasks and story contents on this metacognitive behavior. Future work, however, should explore the effects of both having and not having the opportunity to look back in text. Another limitation is that different readers were assigned to each of the three Response Tasks; future studies should explore the effects of generating vs. evaluating as a repeated measure. Finally, all participants in this study read the same texts, written at the 5th grade level; future studies should use grade-appropriate texts for each reader.

Conclusion

The most noteworthy results of this study were the strong effects of response task and logical form on students' metacognitive behavior when reading texts that require reasoning. Our observations of this behavior have implications at three levels: 1) the understanding of logical reasoning with indeterminate forms, 2) the development of metacognitive behaviors in text processing, and 3) the educational uses of asking readers to evaluate conclusions modeled in text. The latter two are of particular relevance to reading at the secondary level.

Regarding logical reasoning with indeterminate forms, these results demonstrate that we cannot assume people process such forms either heuristically or analytically based solely on their conclusions. Readers assigned to the Invalid task faced a difficult undertaking—resisting the impulse to agree with examples of the most common (though logically invalid) responses to AC and DA forms, even when these responses were explicitly modeled in text. Based only on their conclusions, an assumption of heuristic processing would be likely, since they usually agreed with the invalid inferences. But the high rates of looking back in response to these forms suggest that whatever readers were using, it was unlikely to be the fast, automatic, minimally effortful experiential processing system described by Klaczynski and Daniel (2005). Although this measure of metacognitive behavior cannot tell us for certain what kind of processing readers used, the results suggest that with difficult reasoning problems, readers who were unsuccessful at arriving at logically valid conclusions may nevertheless have attempted to use analytic processing. Future work should attempt to identify other methods of exploring the cognitive processes people use while reasoning with indeterminate forms.

Regarding the development of metacognitive behaviors in text processing, these results illustrate increased proficiency by young adulthood in the use of looking back at previously read text as a general comprehension strategy. But during early and middle adolescence, students' metacognitive behavior is distinctly different when they evaluate, rather than generate, logical conclusions about what they read. This finding argues for further exploration of metalogical change throughout adolescence with other tasks that involve both reasoning and reading.

Finally, the effects of response task on looking back in text have many implications for reading instruction. If asking readers to evaluate conclusions drawn by characters prompts them to engage in a metacognitive behavior that has been demonstrated in other studies to improve comprehension (Burton and Daneman 2007; Kinnunen and Vauras 1995; Walczyk and Taylor 1996), then exploring other reading contexts in which this behavior may improve performance would be worthwhile. While looking back did not always result in better performance on logic questions in this study, the results suggest that looking back is a likely response to particularly difficult text evaluation situations. Since many different texts require critical evaluation of the conclusions they present, examining lookbacks in text may prove useful in exploring adolescent readers' comprehension of advanced reading materials that require critical evaluation.

Acknowledgments The authors wish to thank the staff, parents, and students at P. K. Yonge Developmental Research School, University of Florida, Gainesville, Florida, for their support and participation in this study.

Appendix 1: Excerpts from true, false, and neutral stories

Excerpts from true story: "The Nature Center"

It was a hot, sunny day in the springtime in southern Georgia. On a wooded hillside above the small town of Hayward, a boy and his grandfather were walking from the parking lot to the entrance of the Faltzinger Aquarium and Nature Center................It was for a field trip that the boy, whose name was Joe, and his Grandfather, whose name was Al, came to the Faltzinger Center on this hot day. The children's parents had been invited on the field trip, and many parents were there. But Joe's parents had to work, so Joe had asked his grandfather to come instead.…….........Barb, the Center's naturalist, began the tour by welcoming all the guests.....

Generate Task: (*Denied Antecedent*)

The children were enjoying the birds and squirrels along the path when they heard a noise in the brush nearby. Barb told them to watch quietly. "Hey Grandpa," said Joe, "What if it's a snake?" "If it's a garter snake, it will have stripes," said Barb. Just then, something long and thin slithered across the path and moved away fast. All the children rushed forward to see it, but Joe and his Grandpa didn't get a good look. "That was not a garter snake," said Barb. "Maybe not," said Grandpa, "but *lots of snakes have stripes. It could have been a rainbow snake*, *or a rat snake*, *or even a glass lizard—those have stripes and they look like snakes because they don't have any legs*."

Question: Did the snake on the path have stripes?

Valid Task: (*Denied Antecedent*)

"*It could have been a rainbow snake*, *or a rat snake*, *or even a glass lizard—those have stripes and they look like snakes because they don't have any legs*. So the fact is," Grandpa said, "we still can't tell if that one had stripes or not!"

Question: Was Grandpa right about the stripes on the snake?

Invalid Task: (*Denied Antecedent*)

"*It could have been a rainbow snake*, *or a rat snake*, *or even a glass lizard—those have stripes and they look like snakes because they don't have any legs*. I guess Barb is right, though," Grandpa said. "It must not have had stripes."

Question: Was Grandpa right about the stripes on the snake?

Excerpts from false story: "Twilight Gulch"

.....One hot day, Emily Harris took a trip through the desert in the car with her parents and her brother Pete to visit their cousins' new home. With temperatures reaching 120° in the middle of the day, a desert can be a dangerous place for humans……Later, she wondered if everything they saw that day was a mirage. They didn't realize at first that they were lost. But when they passed a couple of fragrant creosote bushes and found themselves in the town of Twilight Gulch, they knew something was wrong. Everything about that town was downright peculiar. They couldn't find a phone or a motel, and all the buildings were round with no doors……..Before they became thirsty enough to eat a cactus, they met an old man who saw they were strangers and agreed to show them around. His name was Hank. "It's a funny place, all right," Hank said, "but once you get used to our ways you'll get along fine." And they did get along fine, because even though the town was very strange, Emily and her family found that Hank could always be trusted to tell them how things worked.

Generate Task: (*Modus Ponens*)

Pete asked Hank what people did for fun in Twilight Gulch. Hank said he was in a softball league, and his team was called the Twilight Dudes. "Tonight is the last game of the season, so why don't you stay and watch it?" Hank asked. "If we lose this game, we'll be the league champions." Emily and Pete's parents were not inclined to get back on the highway. "It's too late by now, and we're completely exhausted," their father said. So they stayed for the game, and Hank's team lost.

Question: Were the Twilight Dudes the league champions?

Valid Task: (*Modus Ponens*)

So they stayed for the game, and Hank's team lost. "Great!" said Pete. "Now you're the league champions."

Question: Was Pete right about Hank's team?

Invalid Task: (*Modus Ponens*)

So they stayed for the game, and Hank's team lost. "Too bad!" said Pete. "Now you're not the league champions."

Question: Was Pete right about Hank's team?

Excerpts from neutral story: "A Space Mission"

The starship Alliance left Earth on March 10, 2098. Captain Alice Baker led the mission, with a crew of five. It was a dangerous but very important assignment. If it were successful, Earth's people would survive for many generations. But if it failed, all life on Earth faced extinction. Because the mission was so urgent, its actual purpose had been kept a secret. Even Captain Baker did not know for certain where they were going. For over four years the members of the flight crew rested in the deep sleep of suspended animation. While they slept, the Alliance traveled to the Alpha Centauri star system. There, unknown to almost everyone on Earth, a new planet had been discovered. And this small planet held the secret to their survival……… Suddenly, the Alliance's Alert System blasted in Alice Baker's cabin, and she was awake. ………After Alice had eaten and dressed, she was ready to prepare her ship for the crew's awakening. She did not know the purpose of the mission, but all commands for it were programmed into the ship's computer. It was located just outside her cabin in her Command Station. There she could check the Standard Orders for the mission. This screen would tell her what to do next……..

Generate Task: (*Affirmed Consequent*)

Another Standard Order appeared on the main computer screen. It said: "It is important to check the ship's hull to see if it has been damaged by particles of space dust on this voyage. If the ship has been struck by space dust, there will be dents in the hull's outer layer. Scan the complete hull and report back." Alice used the outside video cameras to scan the ship's hull. She noticed many dents in the ship's outer layer. "I don't know if these dents are from space dust or not," she thought. "*There's plenty of debris floating around in Earth's atmosphere that could have caused them when we left*."

Question: Had the ship been struck by space dust?

Valid Task: (*Affirmed Consequent*)

…….."*There's plenty of debris floating around in Earth's atmosphere that could have caused them when we left*." Alice reported back to the computer that it was unknown if the ship's hull had been hit by space dust.

Question: Was Alice correct in her report to the computer?

Invalid Task: (*Affirmed Consequent*)

…….."*There's plenty of debris floating around in Earth's atmosphere that could have caused them when we left*." Alice reported back to the computer that the ship's hull had been hit by space dust.

Question: Was Alice correct in her report to the computer?

Appendix 2: Tables of means and standard deviations for analyses

Table 1 Means and standard deviations for number of lookbacks by grade, response task, and story content

Grade	Condition	False	True	Neutral	Total	N
5th	Generate	5.29 (5.34)	8.71 (4.07)	6.86 (11.36)	20.86 (16.72)	7
	Valid	29.43 (19.38)	22.0 (16.97)	25.71 (15.63)	77.14 (40.76)	7
	Invalid	19.75 (9.93)	34.62 (22.17)	30.50 (22.52)	84.88 (40.37)	8
Total 5th		18.23 (15.72)	22.36 (19.24)	21.45 (19.60)	62.05 (44.02)	22

Table 1 (continued)

Grade	Condition	False	True	Neutral	Total	N
8th	Generate	16.45 (16.23)	20.73 (11.23)	15.64 (12.33)	52.82 (29.34)	11
	Valid	15.10 (12.41)	52.30 (57.97)	22.60 (20.51)	90.00 (87.44)	10
	Invalid	47.56 (12.57)	47.22 (24.78)	65.78 (38.53)	160.56 (45.00)	9
Total 8th		25.33 (20.03)	39.20 (38.27)	33.00 (32.83)	97.53 (72.38)	30
10th	Generate	9.14 (4.10)	21.14 (11.23)	21.14 (35.81)	51.43 (38.13)	7
	Valid	14.43 (15.27)	24.29 (27.72)	10.71 (17.85)	49.43 (52.82)	7
	Invalid	45.86 (28.84)	68.86 (54.86)	46.43 (40.89)	161.14 (120.83)	7
Total 10th		23.14 (24.50)	38.10 (40.86)	26.10 (34.90)	87.33 (92.27)	21
College	Generate	18.25 (13.22)	22.62 (14.52)	24.88 (24.13)	65.75 (48.93)	8
	Valid	37.25 (36.59)	66.62 (53.96)	55.50 (54.61)	159.38 (136.30)	8
	Invalid	34.57 (17.09)	32.71 (19.43)	40.57 (16.95)	119.00 (57.24)	7
Total College		29.83 (25.24)	41.00 (38.47)	40.30 (37.18)	114.52 (95.75)	23
Total	Generate	12.97 (12.50)	18.73 (11.85)	17.18 (21.95)	48.88 (36.98)	33
	Valid	23.63 (23.73)	43.12 (46.78)	28.91 (34.31)	95.66 (94.31)	32
	Invalid	37.06 (20.55)	45.58 (33.97)	46.61 (33.07)	131.77 (74.63)	31
	All Tasks	24.30 (21.63)	35.53 (35.72)	30.59 (32.23)	91.24 (79.30)	96

Table 2 Means and standard deviations for number of lookbacks by grade, response task, and form (Determinate vs. Indeterminate)

Grade	Condition	Determinate	Indeterminate	N
5th	Generate	5.14 (5.87)	5.28 (6.47)	7
	Valid	30.85 (29.43)	25.00 (21.85)	7
	Invalid	22.63 (6.65)	31.87 (19.85)	8
Total 5th		19.68 (19.68)	21.22 (20.31)	22
8th	Generate	16.81 (12.76)	14.18 (9.36)	11
	Valid	24.90 (26.92)	22.60 (22.74)	10
	Invalid	49.67 (34.09)	63.89 (25.65)	9
Total 8th		29.37 (28.22)	31.90 (28.96)	30
10th	Generate	8.57 (4.11)	26.42 (39.76)	7
	Valid	10.85 (15.26)	21.14 (24.88)	7
	Invalid	49.71 (43.47)	68.85 (49.04)	7
Total 10th		23.04 (31.88)	38.80 (43.14)	21
College	Generate	19.87 (15.71)	18.37 (18.47)	8
	Valid	47.00 (40.25)	61.13 (52.78)	8
	Invalid	26.28 (16.14)	54.42 (28.59)	7
Total College		31.26 (28.46)	44.21 (39.97)	23
Total	Generate	13.33 (12.21)	15.90 (21.39)	33
	Valid	28.65 (30.87)	32.44 (35.73)	32
	Invalid	37.41 (30.30)	54.61 (33.51)	31
	All Tasks	26.21 (27.41)	33.91 (34.33)	96

Table 3 Means and standard deviations for number of logically correct responses (out of 4 possible) by grade, response task, and story content

Grade	Condition	False	True	Neutral	Total	N
5th	Generate	2.00 (0.00)	1.83 (0.75)	2.00 (0.63)	5.83 (1.33)	6
	Valid	2.29 (0.76)	3.00 (1.00)	2.86 (1.07)	8.14 (1.77)	7
	Invalid	1.75 (0.46)	2.75 (1.16)	1.88 (0.99)	6.38 (1.99)	8
Total 5th		2.00 (0.55)	2.57 (1.08)	2.24 (0.99)	6.81 (1.94)	21
8th	Generate	1.80 (0.63)	1.90 (0.57)	2.00 (0.67)	5.70 (1.42)	10
	Valid	2.75 (1.04)	3.50 (0.76)	3.50 (0.76)	9.75 (1.75)	8
	Invalid	2.00 (0.00)	2.50 (0.84)	2.00 (0.63)	6.50 (1.22)	7
Total 8th		2.17 (0.816)	2.58 (0.97)	2.50 (0.98)	7.25 (2.33)	25
10th	Generate	2.00 (0.00)	2.14 (1.07)	2.00 (0.58)	6.14 (1.21)	7
	Valid	2.67 (1.51)	3.50 (0.84)	3.33 (0.52)	9.50 (2.17)	6
	Invalid	2.00 (0.00)	3.00 (0.82)	1.86 (0.69)	6.86 (1.34)	7
Total 10th		2.20 (0.83)	2.85 (1.04)	2.35 (0.88)	7.40 (2.08)	20
College	Generate	2.38 (0.92)	3.12 (0.64)	2.62 (0.74)	8.12 (1.13)	8
	Valid	2.87 (0.35)	3.38 (0.74)	3.38 (0.74)	9.62 (1.18)	8
	Invalid	1.86 (0.69)	2.86 (1.07)	2.00 (0.58)	6.71 (1.25)	7
Total College		2.39 (0.78)	3.13 (0.82)	2.70 (0.88)	8.22 (1.65)	23
Total	Generate	2.03 (0.61)	2.26 (0.89)	2.16 (0.69)	6.45 (1.58)	31
	Valid	2.66 (0.94)	3.34 (0.81)	3.28 (0.80)	9.28 (1.75)	29
	Invalid	1.89 (0.42)	2.79 (0.96)	1.93 (0.72)	6.61 (1.45)	28
	All Tasks	2.19 (0.76)	2.78 (0.99)	2.45 (0.93)	7.43 (2.05)	88

Table 4 Means and standard deviations for number of logically correct responses (out of 6 possible) by grade, response task, and form (Determinate vs. Indeterminate)

Grade	Condition	Determinate	Indeterminate	N
5th	Generate	5.16 (1.16)	0.67 (0.82)	6
	Valid	5.00 (0.82)	3.57 (0.98)	7
	Invalid	4.75 (1.04)	1.62 (1.68)	8
Total 5th		4.95 (0.97)	2.00 (1.70)	21
8th	Generate	5.20 (1.22)	0.50 (0.71)	10
	Valid	5.25 (1.16)	4.50 (1.60)	8
	Invalid	5.67 (0.52)	0.83 (1.33)	6
Total 8th		5.33 (1.05)	1.92 (2.20)	25
10th	Generate	5.14 (0.90)	1.14 (1.21)	7
	Valid	5.67 (0.82)	3.83 (1.83)	6
	Invalid	5.57 (0.53)	1.28 (1.38)	7
Total 10th		5.45 (0.76)	2.00 (1.86)	20
College	Generate	5.62 (0.52)	2.50 (0.93)	8
	Valid	5.50 (0.76)	4.13 (0.99)	8
	Invalid	4.85 (1.06)	1.86 (0.90)	7
Total College		5.34 (0.83)	2.86 (1.32)	23

Table 4 (continued)

Grade	Condition	Determinate	Indeterminate	N
Total	Generate	5.29 (0.97)	1.19 (1.19)	31
	Valid	5.34 (0.90)	4.03 (1.35)	29
	Invalid	5.18 (0.90)	1.43 (1.34)	28
	All Tasks	5.27 (0.92)	2.20 (1.82)	88

References

Alessi, S. M., Anderson, T. H., & Goetz, E. T. (1979). An investigation of lookbacks during studying. *Discourse Processes, 2*, 197–212.

Alverman, D. E. (1988). Effects of spontaneous and induced lookbacks on self-perceived high- and low-ability comprehenders. *The Journal of Educational Research, 81*, 325–331.

Baker, L. (2004). Reading comprehension and science inquiry: Metacognitive connections. In E. W. Saul (Ed.), *Crossing borders in literacy and science instruction*: *Perspectives on theory and practice* (pp. 239–257). Newark: International Reading Association, Arlington, VA: NSTA Press.

Burton, C., & Daneman, M. (2007). Compensating for a limited working memory capacity during reading: evidence from eye movements. *Reading Psychology, 28*, 163–186.

Byrnes, J. P., & Overton, W. F. (1986). Reasoning about certainty and uncertainty in concrete, causal, and propositional contexts. *Developmental Psychology, 22*, 793–799.

Chao, S., & Cheng, P. W. (2000). The emergence of inferential rules: the use of pragmatic reasoning schemas by preschoolers. *Cognitive Development, 15*, 39–62.

Daneman, M., & Hannon, B. (2001). Using working memory theory to investigate the construct validity of multiple-choice reading comprehension tests such as the SAT. *Journal of Experimental Psychology: General, 30*, 208–223.

Daniel, D. B., & Klaczynski, P. A. (2006). Developmental and individual differences in conditional reasoning: Effects of logic instructions and alternative antecedents. *Child Development, 77*, 339–354.

Deloache, J. S., Miller, K. F., & Pierroutsakos, S. L. (1998). Reasoning and problem solving. In W. Damon, D. Kuhn, & R. S. Siegler (Eds.), *Handbook of child psychology, Volume 2: Cognition, perception, and language* (pp. 801–850). New York: Wiley.

Evans, J. S. B. T. (2002). Logic and human reasoning: an assessment of the deduction paradigm. *Psychological Bulletin, 128*, 978–996.

Flesch, R., & Lass, A. H. (1996). *The classic guide to better writing: 50th anniversary edition*. New York: Harper Collins Publishers.

Franks, B. A., Mulhern, S. L., & Schillinger, S. M. (1997). Reasoning in a reading context: deductive inferences in basal reading series. *Reading and Writing, 9*, 285–312.

Franks, B. A., Therriault, D. J., Buhr, M. I., Chiang, E. S., Gonzalez, C., Kwon, H., et al. (2011, June). *Thinking aloud*: *Adolescents' logical reasoning on a reading task*. Presented at the annual meeting of the Jean Piaget Society, Berkeley, CA.

Franks, B. A. (1996). Deductive reasoning in narrative contexts: developmental trends and reading skill effects. *Genetic, Social, and General Psychology Monographs, 122*, 76–105.

Franks, B. A. (1997). Deductive reasoning with prose passages: effects of age, inference form, prior knowledge, and reading skill. *International Journal of Behavioral Development, 21*, 501–535.

Garner, R., & Reis, R. (1981). Monitoring and resolving comprehension obstacles: an investigation of spontaneous text lookbacks among upper-grade good and poor comprehenders. *Reading Research Quarterly, 16*, 569–582.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York: Basic Books.

Janveau-Brennan, G., & Markovits, H. (1999). The development of reasoning with causal conditionals. *Developmental Psychology, 35*, 904–911.

Kinnunen, R., & Vauras, M. (1995). Comprehension monitoring and the level of comprehension in high- and low-achieving primary school children's reading. *Learning and Instruction, 5*, 143–165.

Klaczynski, P. A. (2000). Motivated scientific reasoning biases, epistemological beliefs, and theory polarization: a two-process approach to adolescent cognition. *Child Development, 71*, 1347–1366.

Klaczynski, P. A. (2001). Analytic and heuristic processing influences on adolescent reasoning and decision-making. *Child Development, 72*, 844–861.

Klaczynski, P. A., & Daniel, D. B. (2005). Individual differences in conditional reasoning: a dual-process account. *Thinking and Reasoning, 11*, 305–325.

Klaczynski, P. A., Schuneman, M. J., & Daniel, D. B. (2004). Theories of conditional reasoning: a developmental examination of competing hypotheses. *Developmental Psychology, 40*, 559–571.

Markovits, H. (1993). The development of conditional reasoning: a Piagetian reformulation of the theory of mental models. *Merrill-Palmer Quarterly, 39*, 133–160.

Markovits, H. (2000). A mental model analysis of young children's conditional reasoning with meaningful premises. *Thinking and Reasoning, 6*, 335–347.

Markovits, H. (1995). Conditional reasoning with false premises: fantasy and information retrieval. *British Journal of Developmental Psychology, 13*, 1–11.

Markovits, H., & Barrouillet, P. (2002). The development of conditional reasoning: a mental models account. *Developmental Review, 22*, 5–36.

Markovits, H., & Vachon, R. (1989). Reasoning with contrary-to-fact propositions. *Journal of Experimental Child Psychology, 47*, 398–412.

Markovits, H., & Vachon, R. (1990). Conditional reasoning, representation, and level of abstraction. *Developmental Psychology, 26*, 942–951.

Markovits, H., Venet, M., Janveau-Brennan, G., Malfait, N., Pion, N., & Vadeboncoeur, I. (1996). *Child Development, 67*, 2857–2872.

Moshman, D. (1990). The development of metalogical understanding. In W. F. Overton (Ed.), *Reasoning, necessity, and logic: Developmental perspectives* (pp. 205–225). Hillsdale: Erlbaum.

Moshman, D., & Franks, B. A. (1986). Development of the concept of inferential validity. *Child Development, 57*, 153–165.

National Research Council (NRC). (1996). *National science education standards*. Washington, DC: National Academy Press.

Otero, J. (2002). Noticing and fixing difficulties while understanding science texts. In J. Otero, J. A. León, & A. C. Graesser (Eds.), *The psychology of science text comprehension* (pp. 281–307). Mahwah: Lawrence Erlbaum Associates, Publishers.

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading: comparisons to other models. *The Behavioral and Brain Sciences, 26*, 445–526.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-prime user's guide*. Pittsburgh: Psychology Software Tools, Inc.

Simoneau, M., & Markovits, H. (2003). Reasoning with premises that are not empirically true: evidence for the role of inhibition and retrieval. *Developmental Psychology, 39*, 964–975.

Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning*. Mahwah: Erlbaum.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: implications for the rationality debate? *The Behavioral and Brain Sciences, 23*, 645–665.

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology, 95*, 66–73.

Walczyk, J.J. (1995). Testing a compensatory-encoding model. *Reading Research Quarterly, 30*, 396–408.

Walczyk, J. J., & Taylor, R. W. (1996). How do the efficiencies of reading subcomponents relate to looking back in text? *Journal of Educational Psychology, 88*, 537–545.

Zabrucky, K., & Ratner, H. H. (1986). Children's comprehension monitoring and recall of inconsistent stories. *Child Development, 57*, 1401–1418.