

Comprehension skill, inference-making ability, and their relation to knowledge

KATE CAIN and JANE V. OAKHILL
University of Sussex, Brighton, England

MARCIA A. BARNES
*University of Toronto, Toronto, Ontario, Canada
and The Hospital for Sick Children, Toronto, Ontario, Canada*

and

PETER E. BRYANT
University of Oxford, Oxford, England

In this study we investigated the relation between young children's comprehension skill and inference-making ability using a procedure that controlled individual differences in general knowledge (Barnes & Dennis, 1998; Barnes, Dennis, & Haefele-Kalvaitis, 1996). A multi-episode story was read to the children, and their ability to make two types of inference was assessed: coherence inferences, which were essential for adequate comprehension of the text, and elaborative inferences, which enhanced the text representation but which were not crucial to understanding. There was a strong relation between comprehension skill and inference-making ability even when knowledge was equally available to all participants. Subsidiary analyses of the source of inference failures revealed different underlying sources of difficulty for good and poor comprehenders.

Young children's reading comprehension problems have been attributed to deficiencies in a wide range of lower level cognitive processing abilities, such as phonological processing skill (e.g., Shankweiler, 1989), word-decoding facility (e.g., Perfetti, 1985), and vocabulary knowledge (e.g., Beck, Perfetti, & McKeown, 1982; Carroll, 1993). In this article, we focus on a group of children who demonstrate text comprehension difficulties despite proficiency in both word reading and these lower level cognitive skills (see Cain & Oakhill, in press, for a review). The comprehension difficulties of these children must, therefore, arise from impairments in higher level cognitive skills. For example, previous research has shown that children with comprehension difficulties are poor at inference making (e.g., Cain & Oakhill, 1999; Oakhill, 1982, 1984). Such problems have been interpreted within the mental models framework (e.g., Oakhill, 1996). Our findings suggest that poor comprehenders construct incomplete representations of text: They are often able to integrate information

at a local level but are unable to produce a coherent integrated model of the text as a whole. Poor comprehenders' difficulties with inference making are a likely cause of their text-level comprehension problems (Cain & Oakhill, 1999). In the present study, we explored possible sources of poor comprehenders' difficulties with making inferences.

Inference making is regarded as a central component of skilled reading (e.g., Garnham & Oakhill, 1996; Graesser, Singer, & Trabasso, 1994; Singer, 1994; van den Broek, 1994). Although less skilled readers are capable of inferential processing, they do not generate as many inferences as more skilled readers do (e.g., Casteel, 1993; Casteel & Simpson, 1991; Long, Oppy, & Seely, 1997; Oakhill, 1982, 1984; Omanson, Warren, & Trabasso, 1978; Paris & Lindauer, 1976; Paris & Upton, 1976). Thus, it is important to establish which factors limit inference making within such populations.

An inference can be made only when the requisite general knowledge necessary to make that inference is available (e.g., Ackerman, Silver, & Glickman, 1990; Casteel, 1993). Indeed, relevant background knowledge for a passage is a better predictor of fourth graders' ability to generate inferences from and elaborate on that text than is their comprehension skill (Marr & Gormley, 1982). General knowledge differences are, therefore, a potential source of individual differences in inference generation. Using a procedure that ensured that the relevant general knowledge was equally available to all participants prior to inference making, Barnes and colleagues have demonstrated that knowledge availability is not sufficient to ensure adequate

The study reported in this paper was supported by Economic and Social Research Council Grant R000 23 5438 awarded to J.V.O. and P.E.B. The authors gratefully acknowledge the help of the Experimental Psychology Society, whose award to the first author, in the form of a Study Visit Grant, facilitated this work. The authors also thank The Hospital for Sick Children, Toronto, for their kind hospitality on this visit. Finally, thanks to all the staff and pupils from the Brighton and Hove schools who participated in this work. Correspondence should be addressed to K. Cain, who is now at the Department of Psychology, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, England (e-mail: kcain@essex.ac.uk).

inference making in both normally developing children (Barnes, Dennis, & Haefele-Kalvaitis, 1996) and children with the neurodevelopmental disorder of hydrocephalus (Barnes & Dennis, 1998). Given that skilled comprehenders are likely to read more than less skilled comprehenders and, thus, acquire more information from text, it is plausible that their superior inference-making ability may, in part, stem from greater general knowledge.

A primary aim of the present study was to use Barnes's paradigm to determine the extent to which the inference-making problems experienced by children who experience text-comprehension difficulties without neurological disorder may be accounted for by "general knowledge" deficits. To explore this issue, children were first taught a novel knowledge base—a series of facts about an imaginary planet. These facts provided a background for the text that they subsequently read. Individual facts from the knowledge base had to be retrieved and integrated with information in the text in order to generate particular inferences. Answers to the inference questions were considered only if the relevant knowledge-base information was recalled immediately after the story and questions had been completed. This procedure enabled us to investigate inference-making ability when knowledge was equally available to all participants. In addition, the learning and recall trials enabled us to determine whether reading-comprehension ability and skill at drawing inferences were related to differences in the retention of the knowledge base.

Barnes et al. (1996) found that short-term retention of a learned knowledge base was comparable across different age groups but that poor comprehenders with hydrocephalus (a developmental brain pathology) remembered fewer knowledge-base items when retested at the end of the narrative (Barnes & Dennis, 1998). It is not known whether such information is learned or represented differently by poor comprehenders who do not have neurological impairments, the population of interest in the present study. The ability to access such information and integrate it within a model of the text during comprehension may depend on the stability of the information in memory. One index of stability for a knowledge representation is the ability to remember that information over time. We therefore included a delayed memory test for the taught knowledge base, 1 week after the initial experimental session, in order to assess whether comprehension skill and inference-making ability were related to the stability and retention of the knowledge base over time.

Inference-making ability was assessed in the following way. After learning the knowledge base, children were presented with short episodes from a story. Using questions asked after each episode, we assessed their ability to make two types of inference: *coherence* inferences, which are necessary to establish the links between premises in the text, and *elaborative* inferences, which enrich the text representation. Previous authors argue that these two inference types are conceptually distinct and serve different functions in the construction of a text representation (e.g., Garnham, 1982, 1989). In general, readers are sensitive to these

different functions and make a greater number of coherence inferences than elaborative inferences (e.g., Casteel, 1993; Singer, 1994; Whitney, Ritchie, & Clark, 1991).

Previous work has demonstrated that poor comprehenders are poor to generate both types of inference, relative to their skilled peers (e.g., Cain & Oakhill, 1999; Oakhill, 1982, 1984). However, a limitation of this previous work is that the two types of inference have depended on the integration of information from different sources. Generation of a coherence inference required integration of different pieces of information from within the text, whereas generation of an elaborative inference required the reader to integrate information from the text with prior or general knowledge. In the present study, generation of both types of inference depended on the ability to recall the correct textual premise, retrieve information from outside the text (from the taught knowledge base), and integrate these two pieces of information. Thus, we were able to explore whether poor comprehenders were impaired in drawing knowledge-based inferences that served different functions in a text, even when the processing requirements for both inferences were the same. In this study, both types of inference required the integration of a text premise with the knowledge base.

A further aim of the study was to investigate sources of inference failure. Different sources of inference failure have been identified for different populations of children. Failure to recall relevant textual premises is the main source of young children's failure to make coherence inferences (Barnes et al., 1996), but failure to integrate the text premise with the knowledge-base item (when correctly recalled) accounts for the majority of inference failures by older good and poor (garden variety) readers (Barnes & Dennis, 1996). In the present study, different reasons for inference failure were investigated: failure to retrieve the correct premise from the text, failure to recall the relevant item for the knowledge base, failure to integrate the two, or generation of the incorrect inference. These reasons for inference failure are detailed below.

Failure to recall the correct premise from the text may arise because of poor memory for the text per se. When fewer propositions from the story are recalled, a less coherent representation of the text will exist to support recall. Alternatively, the correct premise may not be recalled because there may be failure to encode a particular premise in the first place, either fully or partially. Failure to recall the correct knowledge-base item may occur when the item is available but is, for whatever reason, difficult to retrieve. Knowledge-base items may be less accessible because they may have been encoded less efficiently or retained less precisely. When both items (text premise and knowledge-base item) are available, an inference may not be made because the two pieces of information are not integrated. Finally, children may also fail to generate the correct inference because they make a different one (incorrect inference) or because they are utilizing a different set of criteria for textual cohesion and are not aware that an inference is necessary. Ultimately, failure to generate such inferences, for any of these reasons,

will result in a poorly integrated representation of the text, and comprehension will suffer (Oakhill, 1996).

As stated before, we know that children with adequate word-reading and vocabulary skills but poor text comprehension experience difficulties with inference making and integration (e.g., Cain & Oakhill, 1999). However, although their difficulties have been related to both working memory and metacognitive impairments, the source of inference-making failure is not known. In the present study, we set out to establish the reasons for inference-making differences between skilled and less skilled comprehenders when knowledge was equally available to the two groups. The question of interest here is, Do less skilled comprehenders' difficulties with inference making arise from the same underlying source as those of skilled comprehenders, or do they fail to make as many inferences as their skilled peers because of a different source of difficulty?

In summary, the present study was designed to assess the following issues: (1) to determine whether poor comprehenders have difficulties with two types of knowledge-base inferences that perform different functions in text, (2) to assess the extent to which (general) knowledge deficits affect inference generation, (3) to identify the reasons for inference failure and how they relate to comprehension skill and inference type, and (4) to determine whether less skilled comprehenders experience difficulties with inference generation from texts that they have listened to (since the presentation of the text in the present experiment was auditory).

METHOD

Participants

Two groups of children participated in this study: 7- to 8-year-old skilled comprehenders and less skilled comprehenders. It is now well established that some poor readers' comprehension difficulties stem from poor word-reading skills (e.g., Perfetti, 1985). In this study, we were not interested in generally poor readers, but, rather, we were interested in children who had a specific comprehension deficit in the presence of age-appropriate word-reading skills. Therefore, the skilled and less skilled comprehenders were matched for their ability to read words (both in and out of context) and for chronological age but were selected to differ on a measure of text comprehension. In this way, we aimed to control for the influence of lower level decoding and vocabulary skills on text comprehension (cf. Nation & Snowling, 1998).

There were 13 children in each group, selected using two tests: *The Gates-MacGinitie Primary Two Vocabulary Test* (MacGinitie &

MacGinitie, 1989) and the *Neale Analysis of Reading Ability—Revised British Edition* (Neale, 1989). As stated above, the purpose of the group selection procedure was to select two groups of children with age-appropriate word-reading skills that differed in reading-comprehension ability. We selected these groups by first administering the Gates-MacGinitie test to the entire 7- to 8-year-old population of three junior schools ($n = 163$). This test is group-administered and requires children to select one out of four words to go with the accompanying picture. This test provides a measure of a child's ability to read and understand single words out of context. It was used to screen out "exceptional" readers. These were children who obtained either very low or very high scores and whose reading age (calculated using the Neale Analysis) would be predicted to be either substantially below or above their chronological age. In addition, children whose first language was not English or who had known behavioral, emotional, or language difficulties were excluded from further testing. The remaining children ($n = 79$) were assessed individually using the Neale Analysis.

In the Neale test, children read a series of short stories out loud, and any word reading errors are corrected. They are asked a set of comprehension questions after each story. The passages are graded in difficulty, and testing stops once a prescribed number of reading-accuracy errors has been made. The test provides separate scores for reading accuracy, based on the number of words read correctly, and reading comprehension, based on the number of comprehension questions that the child answers correctly. Performance on the Neale test was used to select and match the two groups (see Table 1 for group characteristics).

The skilled and less skilled comprehenders all obtained age-appropriate reading-accuracy scores and did not differ significantly on this measure [$t(24) < 1.0$]. The skilled group consisted of children whose reading-comprehension scores were at or above those predicted by their reading-accuracy ability, whereas the less skilled group consisted of children whose comprehension scores were depressed relative to their word-reading age. As the values in Table 1 demonstrate, the mean difference between reading accuracy and reading comprehension for the less skilled group was 25 months. In addition, the difference in reading-comprehension age between the skilled and less skilled comprehenders was 30 months [$t(24) = 8.48$, $p < .001$]. The two groups were also matched for chronological age, sight vocabulary (Gates-MacGinitie test), and the number of Neale stories that they had completed (all $ts < 1.0$). The latter measure was necessary to ensure that the difference in comprehension scores did not arise because the less skilled group had read fewer stories and, therefore, obtained lower comprehension scores simply because they had attempted fewer comprehension questions.

Materials and Procedure

All children were tested individually. The materials and procedure were modified from those used by Barnes et al. (1996) and are explained in more detail below. There were three phases to the experiment.

Table 1
Group Characteristics (Means and Standard Deviations)

| Skill Group | Age | | Gates-MacGinitie | | Reading Accuracy | | Reading Comprehension | | Number of Stories | |
|--------------|----------|-----------|------------------|-----------|------------------|-----------|-----------------------|-----------|-------------------|-----------|
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> |
| Less skilled | 8,0 | 3 | 41.3 | 2.01 | 8,10 | 12 | 6,9* | 4 | 3.9 | 0.86 |
| Skilled | 8,1 | 4 | 41.7 | 1.88 | 8,8 | 7 | 9,2 | 12 | 4.1 | 0.76 |

Note—For less skilled and skilled comprehenders, $ns = 13$. Where appropriate, ages are given in years, months, and standard deviations are given in months. The reading accuracy and comprehension scores are the age equivalent scores provided in the Neale test, and the number of stories read refers to the stories that were completed during this assessment. *Less skilled comprehenders obtained significantly ($p < .05$) lower scores than skilled comprehenders.

1. Children were first taught a knowledge base, which comprised 12 facts about an imaginary planet called Gan.

2. When the knowledge base had been taught to criterion (perfect recall), the children were read a six-episode story about this planet. Immediately after they had heard each episode, the children were asked four questions tapping literal and inferential information from that episode.

3. Immediately after the children had heard all six episodes, retention of the knowledge base was retested. It was also retested a week later.

The Knowledge Base

The knowledge base comprised the 12 items used by Barnes et al. (1996) that were relevant for our shortened version of the story. Each item was a piece of information about people, the environment, or a common object that was given a different property on Gan (e.g., "The ponds on Gan are filled with orange juice," "Bears on Gan have bright blue fur," "The turtles on Gan have ice skates attached to their feet") The experimenter read out each item, emphasizing the novel property (underlined above). Acquisition of the knowledge base was then tested using a forced-choice picture-recognition task and a verbal recall task. These two tasks provided an indication of how easily the participants acquired the novel information and served to teach the knowledge base to criterion, as follows.

Forced-choice picture-selection task. This test was administered after the experimenter had read out the knowledge base, as described above. There were 12 trials, one for each item in the knowledge base. The children were presented with four pictures. Their task was to choose the picture that corresponded to the state of affairs on Gan. For example, to test recall of the information that "the turtles on Gan have ice skates attached to their feet," the child's task was to point to the picture that represented the characteristics of the ponds on Gan. The three distractors were (1) the true state of affairs on Earth (turtles without skates), (2) property other than the one ascribed to the object on Gan (turtles wearing roller skates), and (3) the Ganian property ascribed to another object (ducks with ice skates).

Verbal recall task. This test was administered immediately after the picture-selection task described above. There were 12 specific questions to test memory for each item in the knowledge base (e.g., "What are the turtles on Gan like?")

In both tasks, wrong answers were corrected immediately and retested later, after the complete set of items had been presented. Thus, only items that were recalled incorrectly were presented more than once.

The Story Episodes and Questions

After the verbal recall test, the experimenter presented the story episodes. There were six episodes, selected from the original 10 episodes used by Barnes et al. (1996). The main criterion for inclusion of an episode was continuity of story line. Some vocabulary items were changed to British English terms and minor modifications were necessary to ensure that the plot was coherent. Each episode was 142–169 words in length. There were four questions associated with each episode to assess different aspects of text comprehension. These are described below. An example of the story and the four question types is provided in Table 2.

Question types. There were four types of questions: literal content, simile comprehension, coherence inferences, and elaborative inferences.

Literal content. These questions assessed memory for information given literally in the text.

Simile comprehension. For each episode, there was a question that required an explanation of a novel simile that appeared in that episode. General knowledge had to be integrated with information given explicitly in the text, in order to understand the similes.

Coherence inferences. These inferences were necessary to maintain story coherence. Although individuals may differ in their criteria for coherence, this type of inference was regarded as essential

Table 2
Example of Story Episode and Questions

| Episode |
|--|
| The sun was going down and it was getting very cold indeed. Dack and Tane took their coats out of their bags and put them on. Their coats were made of bear's fur. They felt much warmer. Before long the path was icy and slippery. Dack and Tane kept falling on the ice. They saw two turtles ahead of them on the path. "I wish I was a turtle," sighed Dack. Tane slipped and fell on top of her rucksack, crushing all the strawberries that they had picked earlier. When Dack tried to help her up, he fell over too. Dack was covered in scrapes and bruises. He was like a boxer who had lost a fight. "Poor Dack," said Tane picking herself up, "you'll feel better tomorrow." She helped Dack up. Then they walked very carefully along the path, holding each other by the hand. |
| Questions |
| 1. What did Dack and Tane take out of their bags? (elaborative inference) |
| 2. What did Dack wish? (coherence inference) |
| 3. What happened when Tane fell down? (literal information) |
| 4. What does "Dack was like a boxer who had lost a fight mean?" (novel simile) |

to establish cohesion, because particular clauses in the story were anomalous unless integrated with information in the knowledge base. For the example in Table 2, "I wish I was a turtle," sighed Dack" can be fully comprehended only if the information from the knowledge base that the turtles on Gan have ice skates attached to their feet is brought to bear on its interpretation.

Elaborative inferences. These inferences were not necessary to maintain textual cohesion, but they elaborated on story information. For the example in Table 2, it is not necessary to make the inference that the bear fur coats that the children put on were blue. However, the addition of such information would create a richer representation of the text.

All of the inference and literal questions remained the same as those used by Barnes et al. (1996), but alterations to the texts meant that one simile was different.

After the knowledge base had been learned to criterion, the story was read out by the experimenter, one episode at a time. The questions were asked immediately after each episode in the order in which the information occurred in that episode. Therefore, order of question type (coherence inferences, elaborative inferences, literal, and simile) varied across the six episodes, but strict counterbalancing of the different types of question was not possible. When a question was answered incorrectly or incompletely, a nonspecific prompt was used to elicit a fuller answer, "tell me more about that." When an inference was not made, a direct question restating the premise information was asked (e.g., "Why did Dack wish he was a turtle?").

Retention of the Knowledge Base

When all of the story episodes and questions had been completed, memory for the knowledge base was tested once more. The children were asked the same questions as those used in the verbal recall task described above, but, obviously, no feedback or retesting occurred. Performance on this task was taken into account in the scoring of the inference comprehension questions: Only the responses to inference questions for which the knowledge-base items were remembered in this poststory test were included in the total scores used in the analysis. Thus, the participants were not penalized for incorrect responses that were dependent on knowledge-base information that they could not recall. Recall of the knowledge base was tested once more, a week later, using the same task.

RESULTS

The first set of analyses reported below assessed acquisition and memory for the novel information presented in the knowledge base. The next set of analyses assessed performance on the comprehension questions. Subsidiary analyses were conducted to investigate reasons for not making inferences.

The Knowledge Base: Acquiring and Remembering Novel Information

Scoring. One point was awarded for each item correctly recalled the first time, two points for items requiring a second trial, three points for three trials, and so on. The score obtained for each task was the sum of the learning trials required until perfect recall was achieved. Therefore, the score obtained reflects ease of learning: A score of 12 denotes perfect recall (in either the picture choice or the verbal recall test); a score of 14 indicates that either one trial was corrected and retested twice or that two trials were corrected and retested once.

Learning the knowledge base. The scores are presented in the first two columns of Table 3. The picture-selection task was administered before the verbal recall task, and, as expected, the scores on the latter task were slightly lower, indicating that fewer items were retested. Scores on both measures suggest that, as intended, both groups acquired the knowledge base with relative ease. However, there was some indication that the skilled comprehenders were quicker to learn the knowledge base than were the less skilled comprehenders [picture, $t(24) = 1.96, p < .062$; verbal, $t(24) = 1.93, p < .067$].

Retention of the knowledge base. The scores for the immediate and delayed tests of recall of the knowledge base are given in Table 3. A two-way analysis of variance (ANOVA; skill group \times test session) revealed a marginal effect of skill group [$F(1,24) = 3.46, p < .08$], and a significant effect of test session [$F(1,24) = 9.75, p < .006$]. There was a significant interaction between the two variables [$F(1,24) = 9.75, p < .006$]. The interaction arose because the two groups obtained comparable scores in the immediate recall test, but the skilled comprehenders demonstrated superior retention of the knowledge base over time.

Comprehension Questions

Scores obtained on the three question types—literal content, similes, and inferences—were analyzed in separate ANOVAs. All responses were scored according to the guidelines in the manual (constructed on the basis of the study by Barnes et al., 1996) by two raters who were blind to the skill group of the child. Consistency between the raters was high. There were no disagreements on responses to literal and coherence inference questions, one disagreement on an elaborative inference question, and nine disagreements on responses to simile questions (which was fewer than 6% of all simile responses). All disagreements were resolved by discussion.

Memory for literal content. Scores were awarded as follows: 3 points for a full response without prompting (for the example in Table 2 “she squashed all the berries in her bag”), 2 points for a full response after the prompt “tell me more about that,” 1 point for a partial response that was not improved when prompted (e.g., “she fell on top of her bag/she squashed all the strawberries”), and 0 points for incorrect answers and “don’t knows.” There were six literal content questions in total (one in each episode); therefore, the maximum possible score was 18. Mean scores are shown in Table 4 (column 1). A t test revealed that the skilled comprehenders recalled more literal information than did the less skilled comprehenders [$t(24) = 2.65, p < .015$].

Similes. The scoring system was similar to that used for the literal content questions. Example responses refer to the simile in the sample text provided in Table 2. Three points were awarded for a full interpretation of the simile without prompting (e.g., “he was covered in scrapes and bruises”), 2 points for a full interpretation after prompting, 1 point for a partial response that was not improved when prompted (e.g., “he was hurt”) and for “mixed” responses that included some interpretation and some literal description (e.g., “he’s just been in a fight, looks like he got punched”), and 0 points for wholly incorrect answers and “don’t knows.” A t test between the two groups’ scores did not reach significance [$t(24) = 1.63, p > .10$].

Inference-making skill. The total number of inferences correctly drawn when the question was first asked or after the prompt for further information (“tell me more

Table 3
The Knowledge Base: Ease of Learning and Retention Scores
(Means and Standard Deviations) for Each Skill Group

| Skill Group | Ease of Learning | | | | Retention | | | |
|--------------|------------------|-----------|---------------|-----------|-----------|-----------|--------------|-----------|
| | Picture Test | | Verbal Recall | | Immediate | | 1-Week Delay | |
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> |
| Less skilled | 14.08 | 2.25 | 13.62 | 2.40 | 11.77 | 0.69 | 10.77* | 1.48 |
| Skilled | 12.77 | 0.83 | 12.31 | 0.48 | 11.85 | 0.37 | 11.85 | 0.37 |

Note—For the picture test and the verbal recall test, 1 point was awarded for each item correctly recalled the first time. Errors were corrected and retested, and 1 point was added for additional trials needed. Perfect recall score was 12. For the immediate and delayed tests of recall, the maximum score was 12. *Less skilled comprehenders obtained significantly ($p < .05$) lower scores than skilled comprehenders.

Table 4
Literal Questions and Similes: Mean Scores
and Standard Deviations for Each Skill Group

| Skill Group | Literal Questions | | Similes | |
|--------------|-------------------|-----------|----------|-----------|
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> |
| Less skilled | 10.38* | 3.52 | 12.31 | 4.05 |
| Skilled | 13.31 | 1.84 | 14.92 | 4.11 |

Note—For literal questions and similes, the maximum score was 18.
 *Less skilled comprehenders obtained significantly ($p < .05$) lower scores than skilled comprehenders.

about that”) was calculated. These totals were then adjusted to take into account an individual’s retention of the knowledge base when tested immediately after the story presentation. Points were awarded on a matched basis. For example, failure to recall the specific information about the turtles on Gan resulted in the exclusion of the response to that question. There were no instances in which a child produced a correct inferential answer but did not recall the appropriate knowledge-base item at the end of the story. The scores entered into the analysis were each individual’s raw score expressed as a proportion of the total possible score for that child, dependent on their memory of the knowledge base. Thus, if a child recalled all six items of the knowledge-base items on which the elaborative inferences could be drawn, their maximum possible raw score was 6; if they only recalled three of these items, their maximum possible raw score was 3.

The proportional scores were analyzed in a two-way ANOVA, with skill group (skilled, less skilled) and inference type (coherence, elaborative) as factors. The group means are reported in Table 5. There was a main effect of skill group because the skilled comprehenders made more inferences in general than the less skilled comprehenders did [$F(1,24) = 12.31, p < .002$]. Although more coherence inferences were drawn than were elaborative inferences, the effect of inference type was not significant [$F(1,24) = 1.06, p > .10$], and there was no interaction between the two factors, skill group and inference type [$F(1,24) < 1.0$]. Because of the small number of items ($n = 6$) for each inference type, effect sizes were calculated. The effect sizes were substantial: 1.3 and 0.96 for coherence and elaborative inferences, respectively.

Reasons for Inference Failure

There are several reasons why an individual may have failed to make an inference, even when he/she was able to recall the knowledge-base information immediately after story presentation. We were able to explore some of these possibilities in subsidiary data analyses, reported next.

Differential memory of the text and the knowledge base. It is possible that the less skilled comprehenders had poorer memory for the text per se. Indeed, in this study, they were poorer at answering literal questions than the skilled comprehenders (cf. Cain & Oakhill, 1999; Oakhill, 1982). Furthermore, Barnes et al. (1996) found that scores on the literal questions were related to

performance on coherence inference questions. It is also possible that differential memory for the knowledge base affected performance, even though responses to the inference questions were conditionalized for immediate knowledge-base recall: The skilled comprehenders demonstrated better retention of the knowledge base over a 7-day period than did the less skilled comprehenders.

Analysis of covariance was used to determine whether these differences in memory for either the text or the knowledge base could account for differential performance on the inference questions. As noted above, the inference scores were conditionalized for immediate recall of the knowledge base. Therefore, this analysis constitutes a very stringent test of the hypothesis that differences between good and poor comprehenders’ inference-making skill are not simply the result of knowledge differences. Both indicators of differential memory—literal scores and delayed recall of the knowledge base—were entered as covariates to control for a general trend toward differences in memory. The main effect of skill group was significant [$F(1,22) = 4.45, p < .05$].

Stages in the inference-making process. In the introduction, we identified different stages of the inference-making process where difficulties might arise. *Incorrect premise recall* is one type of information retrieval error, which occurs when an individual fails to retrieve the relevant premise from the text. Individuals may either forget the relevant premise or retrieve the incorrect premise from the story. Another type of information retrieval error is a failure to recall the correct item from the knowledge base. This source of inference failure has already been addressed in the analysis of the proportional (adjusted) scores. *Integration failures* occur when children fail to integrate the two relevant pieces of information when retrieved. In addition, individuals may generate an *incorrect inference* in order to make sense of the text and establish coherence by, for instance, integrating the premise information with other (possibly real-world) knowledge. For instance, in one episode, the children get to a high fence. The children put their shoes on and “they flew across the fence, landing gently on the other side.” To fully understand this part of the story, a coherence inference must be made to integrate the textual premise with the knowledge-base information that the shoes on Gan have wings. However, when asked “What did Dack and Tane do at the fence?” one participant responded, “they jumped

Table 5
Inference Questions: Mean Proportional Scores
and Standard Deviations for Each Skill Group
for the Two Types of Inference

| Skill Group | Type of Inference | | | |
|--------------|-------------------|-----------|-------------|-----------|
| | Coherence | | Elaborative | |
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> |
| Less skilled | .359* | .245 | .322* | .228 |
| Skilled | .627 | .168 | .556 | .255 |

*Less skilled comprehenders obtained significantly ($p < .05$) lower scores than skilled comprehenders.

over." Such errors were classified as incorrect inferences and agreed by both markers. The percentages of incorrect coherence inferences were 1.3% and 2.5% for skilled and less skilled comprehenders, respectively. The percentages of incorrect elaborative inferences were 2.5% and 10.3% for skilled and less skilled comprehenders, respectively.

The different error types are not logically independent, so the data were explored in the following way. First, we conducted a two-way ANOVA (skill group \times inference type) on the incorrect inference responses. The analysis of these errors takes into account the number of inference failures made for each type of inference. So, for instance, if there were three coherence failures, and all three could be attributed to generating an incorrect inference, the proportion score for "inference failure" entered into the analysis would be 1.0. If only two of the errors were attributable to incorrect inference generation, the proportion would be .667.

Very few errors could be attributed to an incorrect inference, and an analysis of these errors showed no significant effect of skill group [$F(1,24) = 1.32, p > .10$]. Thus, there was no evidence to suggest that less skilled comprehenders gained lower inference scores because they were generating nontarget inferences. However, a greater proportion of elaborative inference failures, relative to coherence inference failures, could be attributed to incorrect inference generation [$F(1,24) = 4.70, p < .05$]. Incorrect inference generation accounted for 3% of coherence inference failures for both groups. This type of error accounted for 7% of elaborative inference failures made by skilled comprehenders and 15% of less skilled comprehenders' elaborative failures. The interaction between the two factors was not significant [$F(1,24) < 1.0$].

To explore the contribution of incorrect premise recall and integration failure to inference-making difficulties, we calculated the proportion of remaining errors (not including knowledge-base retrieval difficulties or incorrect inference generation) that involved a failure to recall the correct premise from the text. These scores, shown in Table 6, were entered into a two-way ANOVA, with skill group and inference type as factors. The analysis revealed a main effect of skill group [$F(1,24) = 5.33, p < .035$], because a

Table 6
Inference Failures: Mean Proportional Scores and Standard Deviations of Remaining Inference Failures That Could Be Attributed to Incorrect Premise Recall Once Knowledge-Base Failure and Incorrect Inference Generation Had Been Excluded for Each Skill Group for Both Types of Inference

| Skill Group | Type of Inference | | | |
|--------------|-------------------|-----------|-------------|-----------|
| | Coherence | | Elaborative | |
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> |
| Less skilled | .368* | .345 | .526 | .285 |
| Skilled | .103 | .199 | .415 | .368 |

*Less skilled comprehenders obtained significantly ($p < .05$) higher scores than skilled comprehenders.

Table 7
Inferences Made With Direct Questions: Mean Proportional Scores and Standard Deviations, Adjusted to Take Recall of Knowledge Base Into Account, for Each Skill Group for Both Types of Inference

| Skill Group | Type of Inference | | | |
|--------------|-------------------|-----------|-------------|-----------|
| | Coherence | | Elaborative | |
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> |
| Less skilled | .705 | .282 | .751 | .191 |
| Skilled | .974 | .063 | .962 | .073 |

greater proportion of less skilled comprehenders' inference failures could be attributed to a failure to recall the relevant premise from the text, relative to the skilled group's errors. There was also a main effect of inference type [$F(1,24) = 7.08, p < .015$], because this source of difficulty was more common for elaborative inference failures than for coherence ones. The interaction between the two factors was not significant [$F(1,24) = 0.77, p > .10$].

Recognition of the need to make an inference. Skilled comprehenders may generate more inferences than do less skilled comprehenders because they regularly monitor their comprehension and see the need to make inferences to fill in missing details. Barnes et al. (1996) found that children were able to answer direct questions that required an inference (e.g., "Why did Dack wish that he was a turtle?") even when they had not made that inference when originally asked. We incorporated these direct questions into our comprehension questioning procedure in order to determine whether less skilled comprehenders were able to generate inferences when explicitly required to do so. The means in Table 7 represent the total number of inferences made, collated from when the question was first asked, with prompts (if needed) and with direct questions (if needed). Because these data are not independent of those reported earlier, they were not subjected to statistical analysis. Nevertheless, the mean scores are revealing: The skilled comprehenders were now performing at ceiling, and, more interestingly, the less skilled comprehenders' performance improved greatly.

DISCUSSION

The primary aim of the present study was to investigate the relation between reading-comprehension skill and the ability to draw inferences when knowledge was available. To make an inference, information from the text and a taught knowledge base had to be recalled and integrated. When inferences were not made, we explored whether there was a common source of difficulty for good and poor comprehenders and a common source of difficulty for the two inference types, coherence and elaborative. As stated in the introduction, less skilled comprehenders may simply have poorer memory of the information necessary for inference generation. An inference may not be made because of a failure to recall the knowledge-base item or the correct premise from the text

or because of a failure to integrate the two. In addition, children may generate a different inference to that intended, or they may not draw an inference because they are not aware that one is necessary. Inference failure, for any of these reasons, will result in a less detailed and integrated model of the text. We summarize and discuss the results as they relate to these points, in turn.

The procedure was designed to ensure that all children could learn the knowledge base from which the inferences could be drawn with relative ease. There were ceiling effects in both immediate and delayed recall of the knowledge base, but these were a necessary consequence of the task requirements (to learn the knowledge base to criterion and to retain it). The results suggest that children with good comprehension skills may find it easier to acquire new knowledge and are also able to construct more stable representations of newly taught knowledge than less skilled comprehenders, even when their short-term retention does not differ markedly. However, the present study was not designed to assess the acquisition and retention of knowledge, although these are issues that warrant further investigation. The skilled comprehenders' superior recall of the knowledge base may also have been aided by their better memory for the story. Specifically, they may have constructed a more integrated and embellished representation of the text, which may have served to strengthen their memory of the knowledge base, such that knowledge-base items may have been available as an integral part of the story rather than a list of discrete facts.

The differences that existed in availability of the knowledge base were taken into account when inferring skill was assessed. The less skilled comprehenders generated significantly fewer inferences than the skilled comprehenders did. The effect sizes revealed that the group differences in inference-making skill were substantial (Cohen, 1988). Thus, even when they had the requisite knowledge-base information from which to generate an inference, the less skilled comprehenders did not make these inferences as readily as their skilled peers did. Knowledge availability is therefore not a sufficient condition for inferring, and we can rule out lack of knowledge as a primary source of poor comprehenders' inference-making difficulties. Furthermore, analysis of covariance demonstrated that the skilled comprehenders' superior inference-making skills were not simply due to differences in their memory for either the text or the knowledge base over time. As stated above, knowledge for the story may serve to strengthen memory for the knowledge base. Thus, the inclusion of delayed knowledge-base recall in our analysis provides a particularly strong test of the hypothesis that inference-making differences could be attributed to differential memory. The group difference remained even when both indicators of differential memory were entered into the analysis.

There are two qualifications to the conclusion that differential memory did not affect inference-making performance. First, the skilled comprehenders demonstrated ceiling

performance on the test of long-term retention of the knowledge base. This restricted range of scores may have limited the explanatory power of the retention measures. Second, it may be that these variables are not suitable controls for knowledge stability or accessibility. They do not assess the degree of integration between the items acquired in the new knowledge base, nor do they assess the speed or efficiency of access of this new information. Barnes et al. (1996) found that easily accessible knowledge was more likely to be used in inferring than was knowledge that took longer to retrieve. We were not able to test accessibility of knowledge in the present study, but it is plausible that poor comprehenders were restricted by the accessibility of information in the taught knowledge base to a greater degree than were good comprehenders. Measures of knowledge accessibility should be included in follow-up studies.

In this study, we found that literal memory for the text did not account for group differences in inference making. In contrast, Barnes et al. (1996) found that literal memory for the text in general was a significant predictor of coherence inference-making ability within a population of 6- to 15-year-olds. Our present finding is supported by a study conducted by Omanson et al. (1978), who demonstrated that age-related gains in general memory capacity could not wholly account for developmental improvements in inference making. The discrepancies between different studies suggest that it may be necessary to explore different aspects of literal memory in more detail in future work, such as the quality (i.e., detail) of the literal recall, as well as the quantity.

A common source of inference difficulty for the less skilled comprehenders in the present study was a failure to retrieve the relevant textual premise. The primary source of skilled comprehenders' inference failures occurred at a different stage in the comprehension process. Often, they recalled both the relevant textual premise and the knowledge-base item but failed to integrate the two. Recall of the incorrect premise suggests that the less skilled comprehenders experienced difficulty in selecting the relevant information on which the inference should be based. Despite the finding that a higher proportion of the good comprehenders' failures can be attributed to integration failures, relative to the less skilled group, it is certainly not the case that less skilled comprehenders do not experience integration failures. Rather, the less skilled comprehenders' difficulties arose at an earlier stage in the inference-making process: They often failed to recall the information that had to be integrated to generate the inference.

A small proportion of inference failures in both groups could be attributed to generating the incorrect inference. This type of error was more common for elaborative inferences than for coherence ones. Generation of the wrong inference is an indication that the reader (or listener) is poor at selecting the relevant information from the text and from his/her general knowledge (in this instance, the taught knowledge base). There were very few instances of incor-

rect inference generation and, thus, no indication that the less skilled comprehenders gained lower inference scores because they were generating nontarget inferences.

The groups did not generate a significantly greater number of coherence inferences than elaborative ones, although there was a trend in that direction, suggesting that both groups of children were sensitive to the need to maintain textual coherence by making necessary inferences. Previous studies have explored coherence and elaborative inferences and report a difference between these two types of inferences (e.g., Casteel & Simpson, 1991). However, none of these studies have required the integration of information from both text and a knowledge base for both types of inferences: Coherence inferences can often be generated from information provided in the text alone by, for example, integrating two propositions. This difference in processing requirements may be one reason for the smaller than expected difference between coherence and elaborative inferences that was found in the present study. Using the same texts, Barnes et al. (1996) found the smallest difference between the two inference types for 8- to 9-year-olds. Thus, our apparent lack of differentiation between these types for 7- to 8-year-olds may be an indication of developmental differences. It could be that, once word reading has become fairly fluent and text-comprehension skills are beginning to develop independently, children are still learning which inferences are necessary and which are merely elaborative, and they are having to adjust their standards accordingly.

Interestingly, different errors were associated with the two inference types. Incorrect inference generation and retrieval of incorrect textual premises were more common sources of elaborative inference than coherence inference failures. These different patterns of error indicate that, even though there was no overall advantage for coherence over elaborative inferences, the children were less aware of which information was relevant for elaborative inference generation than for coherence inference generation.

The less skilled comprehenders were not significantly impaired in their ability to interpret novel similes. This finding is somewhat surprising, because many of the cognitive processes necessary to make inferences are likely to be involved in simile interpretation—namely, application of (general) knowledge from outside the text to a textual premise. Although similes are a type of figurative expression, they offer clues to their nonliteral interpretation in much the same way that direct questions indicate that an inference is required. It may be that comprehension of a different form of figurative language, which requires a greater degree of inferential processing, such as unfamiliar or novel idiomatic phrases, may be more strongly related to comprehension skill—a hypothesis that we are currently pursuing.

To summarize, the relation between comprehension skill and inference making is now well established. In the present study, we demonstrated that children with comprehension difficulties are deficient at making inferences that require integration of textual premises with a taught

knowledge base, even when those inferences are necessary for comprehension. In addition, we demonstrated that less skilled comprehenders' difficulties with inference making are not just restricted to reading situations but are apparent in tasks involving listening comprehension as well. Less skilled comprehenders' difficulties with inference making were not wholly accounted for by memory for the text or information outside of the text that was essential for inference generation. An analysis of errors revealed that a more likely source of inference-making difficulty for this group was an inability to select the information relevant to making the inference.

REFERENCES

- ACKERMAN, B. P., SILVER, D., & GLICKMAN, I. (1990). Concept availability in the causal inferences of children and adults. *Child Development*, *61*, 230-246.
- BARNES, M. A., & DENNIS, M. (1996). Reading comprehension deficits arise from diverse sources: Evidence from readers with and without developmental brain pathology. In C. Cornoldi & J. V. Oakhill (Eds.), *Reading comprehension difficulties: Processes and interventions* (pp. 251-278). Hillsdale, NJ: Erlbaum.
- BARNES, M. A., & DENNIS, M. (1998). Discourse after early-onset hydrocephalus: Core deficits in children of average intelligence. *Brain & Language*, *61*, 309-334.
- BARNES, M. A., DENNIS, M., & HAEFELE-KALVAITIS, J. (1996). The effects of knowledge availability and knowledge accessibility on coherence and elaborative inferencing in children from six to fifteen years of age. *Journal of Experimental Child Psychology*, *61*, 216-241.
- BECK, I. L., PERFETTI, C. A., & MCKEOWN, M. G. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology*, *74*, 506-521.
- CAIN, K., & OAKHILL, J. V. (1999). Inference making and its relation to comprehension failure. *Reading & Writing*, *11*, 489-503.
- CAIN, K., & OAKHILL, J. V. (in press). Reading comprehension difficulties. In P. E. Bryant & T. Nunes (Eds.), *International handbook of children's reading*. Dordrecht: Kluwer.
- CARROLL, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.
- CASTEEL, M. A. (1993). Effects of inference necessity and reading goal on children's inference generation. *Developmental Psychology*, *29*, 346-357.
- CASTEEL, M. A., & SIMPSON, G. B. (1991). Textual coherence and the development of inferential generation skills. *Journal of Research in Reading*, *14*, 116-129.
- COHEN, J. (1988). *Statistical power analysis for behavioral sciences* (2nd ed.). New York: Academic Press.
- GARNHAM, A. (1982). Testing psychological theories about inference making. *Memory & Cognition*, *10*, 341-349.
- GARNHAM, A. (1989). Inference in language understanding: What, when, why and how. In R. Dietrich & C. F. Graumann (Eds.), *Language processing in social context* (pp. 153-172). Amsterdam: North-Holland.
- GARNHAM, A., & OAKHILL, J. V. (1996). The mental models theory of language comprehension. In B. K. Britton & A. C. Graesser (Eds.), *Models of understanding text* (pp. 313-339). Hillsdale, NJ: Erlbaum.
- GRAESSER, A. C., SINGER, M., & TRABASSO, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*, 371-395.
- LONG, D. L., OPPY, B. J., & SEELY, M. R. (1997). Individual differences in readers' sentence- and text-level representations. *Journal of Memory & Language*, *36*, 129-145.
- MACGINITIE, W. H., & MACGINITIE, R. K. (1989). *Gates-MacGinitie reading tests*. Chicago: Riverside.
- MARR, M. B., & GORMLEY, K. (1982). Children's recall of familiar and unfamiliar text. *Reading Research Quarterly*, *18*, 89-104.
- NATION, K., & SNOWLING, M. J. (1998). Semantic processing and the development of word-recognition skills: Evidence from children with

- reading comprehension difficulties. *Journal of Memory & Language*, **39**, 85-101.
- NEALE, M. D. (1989). *The Neale analysis of reading ability—Revised British edition*. Windsor: NFER-Nelson.
- OAKHILL, J. V. (1982). Constructive processes in skilled and less-skilled comprehenders' memory for sentences. *British Journal of Psychology*, **73**, 13-20.
- OAKHILL, J. V. (1984). Inferential and memory skills in children's comprehension of stories. *British Journal of Educational Psychology*, **54**, 31-39.
- OAKHILL, J. V. (1996). Mental models in children's text comprehension. In J. V. Oakhill & A. Garnham (Eds.), *Mental models in cognitive science: Essays in honour of Phil Johnson-Laird* (pp. 77-94). Hove, U.K.: Psychology Press.
- OMANSON, R. C., WARREN, W. M., & TRABASSO, T. (1978). Goals, inferential comprehension and recall of stories by children. *Discourse Processes*, **1**, 337-354.
- PARIS, S. G., & LINDAUER, B. K. (1976). The role of inference in children's comprehension and memory for sentences. *Cognitive Psychology*, **8**, 217-227.
- PARIS, S. G., & UPTON, L. R. (1976). Children's memory for inferential relations in prose. *Child Development*, **47**, 660-668.
- PERFETTI, C. A. (1985). *Reading ability*. Oxford: Oxford University Press.
- SHANKWEILER, D. (1989). How problems of comprehension are related to difficulties in decoding. In D. Shankweiler & I. Y. Liberman (Eds.), *Phonology and reading disability: Solving the reading puzzle* (pp. 35-68). Ann Arbor: University of Michigan Press.
- SINGER, M. (1994). Discourse inference processes. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 479-515). San Diego: Academic Press.
- VAN DEN BROEK, P. (1994). Comprehension and memory of narrative texts: Inferences and coherence. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 539-588). San Diego: Academic Press.
- WHITNEY, P., RITCHIE, B. G., & CLARK, M. B. (1991). Working-memory capacity and the use of elaborative inferences in text comprehension. *Discourse Processes*, **14**, 133-145.

(Manuscript received June 4, 1999;
revision accepted for publication May 2, 2001.)