Assessing school-aged children's inference making: the effect of test story format in listening comprehension

Jenny Freed*

University of Manchester

Kate Cain

Lancaster University

Dr Jenny Freed: Ellen Wilkinson Building, School of Psychological Sciences, University of Manchester, Oxford Road, Manchester, M13 9PL, UK email: Jenny.Freed@Manchester.ac.uk; telephone: + 44 (0)161 306 1672; fax: + 44 (0)161 275 3373.

Prof Kate Cain: Department of Psychology, Lancaster University, Lancaster LA1 4YF, UK email: K.Cain@Lancaster.ac.uk; telephone: Tel: +44 (0)1524 593990.

Accepted for publication in International Journal of Language and Communication Disorders

Keywords: Inference, Local coherence, Global coherence, School-aged children

Abstract

Background. Comprehension is critical for classroom learning and educational success. Inferences are integral to good comprehension: Successful comprehension requires the listener to generate local coherence inferences, which involve integrating information between clauses, and global coherence inferences, which involve integrating textual information with background knowledge to infer motivations, themes, etc. A central priority for the diagnosis of comprehension difficulties and our understanding of why these difficulties arise is the development of valid assessment instruments.

Aims. We explored typically developing children's ability to make local and global coherence inferences using a novel assessment of listening comprehension. Our aims were to determine whether children were more likely to make the target inferences when these were asked during story presentation vs. after presentation of the story, and whether there were any age differences between conditions.

Methods & Procedures. Children in Years 3 (n=29) and 5 (n=31) listened to short stories presented either in a segmented format, in which questions to assess local and global coherence inferences were asked at specific points during story presentation, or in a whole format, when all of the questions were asked after the story had been presented.

Outcomes & Results. There was developmental progression between age groups for both types of inference question. Children also scored higher on the global coherence inference questions than the local coherence inference questions. There was a benefit of the segmented format for younger children, particularly for the local inference questions.

Conclusions & Implications. The results suggest that children are more likely to make target inferences if prompted during presentation of the story, and that this format is particularly facilitative for younger children and for local coherence inferences. This has implications for

the design of comprehension assessments as well as for supporting children with comprehension difficulties in the classroom.

Keywords: Inference, Local coherence, Global coherence, School-aged children

What this paper adds

What is already known about this subject?

Comprehension is critical for classroom learning and educational success. Successful comprehension requires the listener to generate local coherence inferences (integrating information between clauses) and global coherence inferences (integrating textual information with background knowledge to infer motivations, themes, etc.).

What this study adds

The present study investigated inference making ability in typically developing children using a novel assessment of listening comprehension that enabled us to compare inference making during passage presentation (text presented in a segmented format with questions asked after each segment) with inference making performance for questions asked after passage presentation (traditional whole story presentation format). There was developmental progression in skills between age groups in both studies for both types of inference question. There was a benefit of asking questions during story presentation for the younger children, particularly for the local inference questions. This has implications for the design of comprehension assessments and for supporting children with comprehension difficulties.

Assessing children's inference making: the effect of text format

Successful comprehension involves extracting the meaning of the text. Good comprehenders do not encode a text verbatim, instead they retrieve the meanings of the individual words, put these together to form sentences, and make links between these individual words and sentences to form a memory-based representation of the text's meaning called a mental model (Johnson-Laird, 1980, Kintsch, 1998). Comprehension is critical for classroom learning and educational success across the curriculum and the societal importance of literacy is recognised in the National Curriculum for England (Department for Education, 2014) where emphasis is placed on both listening and reading comprehension in the programmes of study for Key Stage 1 (age 5-7) and Key Stage 2 (age 7-11).

Approximately 10% of primary school-aged children have unexpectedly poor reading and listening comprehension relative to age appropriate word reading skills (Cain, 2009). These children are often overlooked within their classroom environment due to their adequate ability to read aloud single words and sentences (Nation et al., 2004). Because of its importance for educational attainment, it is imperative that we have a better understanding of why comprehension fails for some during this critical period of schooling, in order to develop appropriate support for those who have difficulties in this fundamental skill. The work reported in this paper tested the utility of a novel method for the assessment of listening comprehension with typically developing children. The broader aim was to be able to better diagnose comprehension difficulties and understand their source. We chose to test this novel assessment method in a listening comprehension format to remove the influence of word reading ability, which would be an additional consideration in a reading comprehension task.

Our focus is inference making. Inferences are integral to good comprehension: a listener will make inferences to expand on information within a text, connecting information and ideas both within the text and with previous world knowledge, as they construct the mental model of its meaning (Garnham, 2010, Cain and Oakhill, 1999). There are various ways to describe inferences but there is lack of consensus between researchers on how they should be categorised (Graesser et al., 1994). For the purpose of this study, we distinguish between two types of inference that are necessary to ensure coherence of the mental model that comprehenders construct: local and global coherence inferences (Graesser et al., 1994). Local coherence inferences are made when listeners use information provided in the text to integrate the meanings of clauses and sentences within the text. Sometimes these are cued by pronouns, or can be made by appreciating the link between synonyms or category exemplars, for example *"He noticed the <u>wild birds</u> swimming on the lake. The <u>ducks</u> were moving towards the island."* (underlining of the words that signals the local coherence inference). In contrast, inferences necessary for global coherence are not typically cued in the same way and are more dependent on the integration of information in the text with the child's background knowledge (Cain and Oakhill, 1999). This can involve inferring motivations or establishing a theme or character identity within a text, for example inferring that the action takes place at circus from key terms such as *'the big top', 'clowns',* and *'trapeze'*. This distinction between global and local coherence inferences has also been adopted in other studies with children (e.g. Cain and Oakhill, 2014, Currie and Cain, 2015) and adults (e.g. Long et al., 1994, Long and Chong, 2001) and have previously been referred to as gap-filling and text-connecting/bridging inferences respectively (Cain and Oakhill, 1999, Graesser et al., 1994, Baker and Stein, 1981).

The notion of coherence is central to good comprehension. van den Broek and colleagues (van den Broek et al., 2001a, van den Broek et al., 1995) have developed the standards of coherence framework to identify the implicit and explicit criteria used by comprehenders to ensure adequate comprehension of text. Comprehenders will adjust their standard of coherence depending on the context or demands of the comprehension task (Kendeou, 2014). An individual's standard of coherence will guide the type and number of inferences that are drawn, such that a good comprehender will engage in an appropriate level of inference generation to maintain adequate coherence (van den Broek et al., 2001a). Thus, standards of coherence have a direct impact on the level of comprehension that is achieved and the mental representation that is constructed (van den Broek et al., 2001a).

Children are aware of the need to generate both local and global coherence inferences from an early age (Ackerman, 1986, Lynch et al., 2008) but they do not always make sufficient coherence inferences to ensure adequate comprehension. The ability to generate both types of coherence inference improves significantly between 4 to 10 years (Currie and Cain, 2015, Lynch et al., 2008, Ackerman, 1988). Differences in background knowledge do not fully account for these developmental differences: when differences in knowledge are controlled, younger children and poor comprehenders continue to make fewer inferences than older children and good comprehenders (Barnes et al., 1996, Cain et al., 2001). An alternative explanation for these developmental improvements proposes a developmental change in the standards of coherence children apply to text (Kendeou et al., 2014): young children have been characterised as processing text in a piecemeal manner (Schmidt and Paris, 1983).

Children with poor reading and language comprehension have problems with inference making (Cain and Oakhill, 1999, Bishop and Adams, 1992). A central priority for

the diagnosis of comprehension difficulties and our understanding of why these difficulties arise is the development of valid assessment instruments. Standardised measures of passage comprehension (e.g. NARA; Neale, 1997, YARC; Snowling et al., 2009, WIAT; Wechsler, 2005, understanding spoken paragraphs on CELF; Semel et al., 2006) use a 'whole' presentation format where the child first listens to or reads a whole passage and subsequently answers a series of questions about the passage. The same procedure is typically adopted in studies of children's inference making (Cain and Oakhill, 1999, Bishop and Adams, 1992). This method provides an overall indication of comprehension ability and is useful for identifying text comprehension difficulties in children with language and communication problems, as well as identifying poor comprehenders in mainstream classrooms (Bishop and Adams, 1992, Cain et al., 2000). However, this format assesses inference making after the text has been presented. Language comprehension is a dynamic process and inferences need to be made as a story unfolds in order to construct a coherent meaning-based representation. Thus, this traditional paradigm for comprehension testing might not provide an accurate estimate of children's comprehension or inference making *potential* and, therefore, does not provide insights into the best type of support for children with comprehension (and inference making) difficulties.

In the present study, we investigated a novel approach to assessment, designed to shed light on inference making potential and reasons for difficulties with inference making. Our novel approach was to compare inference making ability for spoken passages presented in a segmented format with the traditional 'whole' format. This dynamic approach to the assessment of comprehension utilises facilitative techniques currently used by speech and language therapists developing assessment procedures with children who have speech and language disorders in an educational setting (Camilleri and Law, 2007, Hasson and Joffe,

2007, Hasson et al., 2012). This method of comprehension assessment has the potential to inform teachers and speech and language therapists about testing and support of comprehension skills.

By comparing performance in the two conditions, we can identify whether children are better able to make inferences when the memory demands are minimised and the relevant content is readily available (by asking the inference tapping question immediately after the relevant text in the segmented format). Thus, better comprehension ability for passages presented in a segmented format gives an indication about whether any difficulties arise from a tendency to process text in a piecemeal fashion rather than integrating information into the mental model as the text unfolds to ensure high standards of coherence. It may be the case that individuals currently identified as having comprehension difficulties are capable of making sufficient inferences to maintain coherence as they process text , but that existing measures are not sufficiently sensitive to capture this because comprehension questions are traditionally asked at the end of the task (Carlson et al., 2014). Therefore the segmented format should allow comparisons between inferences that are generated *during* the story and those that are identified *after* the material has been presented. The present study will include both global and local coherence inferences and explore whether segmented vs. whole presentation has a different effect on each.

The research base to date is not clear about whether segmented text will benefit comprehension. Asking questions during text presentation has parallels with the 'think aloud' technique, which has been used to study children's comprehension. The think aloud technique requires children to explain, after listening to each sentence of a story, what they have understood so far. Thus, it explicitly segments the processing of text. Laing and Kamhi

(2002) found that both good and poor comprehenders' performance improved when using 'think-aloud' techniques compared to traditional testing methods. Children's active engagement with the text in the 'think-aloud' condition may have encouraged higher standards of coherence and the creation of a situation model, enabling even poorer comprehenders to create the inferences so integral to adequate comprehension of the text. However, the 'think-aloud' technique is cognitively demanding and may preclude useful data from children with language and communication difficulties such as expressive language difficulties. Because our broader aim was to develop an assessment suitable for clinical applications, we decided to compare whole texts with a segmented format. The segmented format could be considered less cognitively demanding than think aloud techniques because the children are responding to specific questions during the text rather than providing an open ended narrative of their ongoing understanding.

The segmented format may facilitate language comprehension by supporting the online inferential processes required in the construction of the mental representation of the story (Olson et al., 1985). In this way, segmented text presentation might scaffold the child's comprehension because the questions focus their attention on relevant parts of the text, highlighting specific details and prompting certain links to be made (van den Broek et al., 2001b). According to the standards of coherence framework, segmented presentation should result in better performance *if* the problem is that children fail to automatically make inferences as they listen. There is a lack of research into inference making for listening comprehension tasks. However, given the overlaps between listening and reading comprehension tasks, research from the latter modality provides useful background information for this study.

Carlson et al (2014) found no overall difference in the number of inferences generated by 9-10 year olds with different levels of comprehension skill when asked causal questions during reading. This contrasts with findings from studies that use the traditional approach, assessing comprehension *after* reading is completed (Oakhill, 1984, e.g. Cain and Oakhill, 1999). Those studies find that struggling readers generate fewer inferences than good readers. Carlson et al (2014) suggest that children who struggle with reading comprehension are *capable* of generating inferences and that being asked questions *during* reading may have decreased the task demands involved in the integration of information within the text sufficiently to allow them to generate appropriate inferences. It should be noted that struggling readers in their study still showed some impairment compared to good readers: they generated fewer knowledge-based inferences related to the text compared to their peers. We note, however, that segmented text may not support all aspects of text comprehension for all children. van den Broek et al (2001b) found that questioning during text reading was challenging for primary aged children and resulted in poorer story retells. They propose that it interrupted and diverted their already limited attentional resources and interfered with the construction of a coherent mental model of the text. Thus, segmented text may benefit only children with weaker comprehension or memory and attentional skills.

Aims of study

The present study was designed to investigate listening comprehension skills in typically developing children, between seven and ten years old, using a novel assessment that compared whole and segmented presentation of spoken passages in order to compare inference making ability during and after text presentation. Comprehension questions assessed the ability to establish both global and local coherence through inference making.

Our aims were to determine whether segmented presentation benefited inference making and whether the effects were consistent across the primary school age range or specific to younger children. The present study also allowed us to explore developmental progression in comprehension skills of a typical population. If the segmented presentation format has a facilitative effect by reducing processing demands and enhancing children's comprehension of the text, we should find an advantage for this presentation format compared with the traditional whole presentation. If our materials are sensitive to developmental change, the older children should attain higher scores for both global and local coherences.

Method

Participants

Sixty children (29 boys) in Years 3 and 5 from mainstream primary schools in a small town in the north west of England took part in the study: 29 (14 boys) were in Year 3 classrooms and were aged 7 to 8 years; and 31 (15 boys) were in Year 5 classrooms and were aged 9 to 10 years. Children who were unable to speak or understand English were excluded as well as any children with known learning, speech or language impairments or children with severe physical or sensory abnormalities that inhibit movement or prevent them from being able to hear or respond to the stimuli. Children with English as a second language were included providing they were fluent in spoken English (decisions on inclusion/exclusion were made following discussion with the class teacher erring on a strict criterion of fluency). Consent was obtained from headteachers, parents/caregivers and the children.

Measures

Comprehension task. Test materials consisted of six short stories similar to those used in school-based comprehension tasks. The stories used were kindly made available by the Language and Reading Research Consortium (LARRC, 2015) for use in this study. Materials

were of appropriate interest (e.g. buying a pet, a birthday party, playing football) and language level for children across the age ranges included and did not contain any sensitive or distressing topics. There were eight comprehension questions related to each story that assessed ability to establish local and global coherence (4 questions for each type of inference). The questions were either presented at intervals during the story (segmented version) or at the end of the story (whole version). The order of the questions followed the order of information presented in the story.

The stories were adapted to be used in this task. For the segmented presentation format, the first author split each story into five segments aiming to keep the segments of a similar length whilst maintaining story coherence. Each segment was followed by one or two questions. Another researcher independently split the stories and comparisons were made between the two versions. There were only two discrepancies in terms of where the stories should be split and these were resolved by discussion. Minor modifications to vocabulary and grammar were also made to make the stories suitable for British English participants (e.g. 'supermarket' instead of 'grocery store'). For each story there was a whole and segmented version (these were exactly the same in terms of story content and questions, only the format of administration differed). The six stories were then grouped into two sets of three stories (Set A and Set B). Each child completed both sets of stories, one in whole format and one in segmented format. The set (A or B) that was completed in each format and the order that they were completed in (whole/segmented, A/B) was counterbalanced independently for both Year 3 and Year 5 children (see Appendix A for example story and questions). The average number of words per story was 157. Children were not presented with any pictures to accompany the story. This was to prevent them from attempting to use clues in the images to answer the questions.

Procedure

Children were seen individually by the researcher in a quiet room in school on two separate occasions (a minimum of two days apart). On each testing session (lasting approximately 30 minutes) the participant was presented with one set of stories (Set A/Set B). Each set consisted of three test stories, presented one at a time. In one of the sessions the stories were presented in in whole format and in the other they were presented in segmented format. The procedure (whole or segmented presentation) was explained to children at the beginning of the session and they were informed that they would be asked questions about the story either after the story (whole presentation) or at various points during the story (segmented presentation). Children then completed a practice story to allow them to become familiar with the task format for that session. Stories were pre-recorded and presented using a PowerPoint file on a laptop computer. This ensured that all children received the same input (i.e. to control for pace of delivery, word stress etc.).

Whole presentation procedure. The child listened to the first story. At the end of the story they answered the eight comprehension questions relating to the story. The remaining stories were then completed according to the same procedure.

Segmented presentation procedure. The child listened to the first story in a series of five segments. After each segment of the story the child was asked one or two comprehension questions (eight in total for the story). The remaining stories were then completed according to the same procedure.

The comprehension questions were read out loud by the experimenter. Prompts were used during the questions where necessary to encourage the child to expand on their answer. The prompt consisted of either a repetition of the question or encouragement for the child to be more specific in their answer. The results reported here relate to the child's

first answer (i.e. before prompting). Children were not told whether they had answered questions correctly, but positive encouragement was given throughout. Children's responses were written down verbatim for later scoring. Each inference question was coded as correct (2 points), partially correct (1 point) or incorrect (0 points) such that each story was scored out of a maximum of 8 points for local inferences and 8 points for global inferences (4 questions x 2 points for each), and each set of stories was scored out of a maximum of 24 (8 points x 3 stories) for both local and global inferences. A scoring rubric was developed which provided examples of full, partial and incorrect answers for each comprehension question (see Appendix B for example scoring rubric). All responses were scored independently by the researcher who had administered the comprehension task as well as by a second researcher blind to the aims and hypotheses of the study. Overall there were disagreements on scoring for 126 questions out of 2880 questions (60 children x 6 stories x 8 questions), showing an agreement level of 95.62%. The 126 disagreements related to 20/48 individual questions from 4/6 stories. For instances where there was a discrepancy in scores, this was resolved by discussion between the two researchers who had done the scoring in order to provide a final set of scores to be used for the analysis.

Results

The set (A or B) that was completed in each format and the order that they were completed in (whole/segmented, A/B) was counterbalanced. Prior to the main analysis, t tests confirmed there were no significant differences between scores from the first and second testing sessions (local: $t(59) = .18$, $p = 0.86$, global: $t(59) = -.07$, $p = 0.95$), suggesting that children did not demonstrate any practice effects in the second session having done one format of the task.

Table 1 around here

Table 1 shows the mean scores (out of 24) for local and global inference questions in whole and segmented presentation format for Year 3 and Year 5 children. Skew and kurtosis confirmed an acceptable data distribution (all values within +/- 1). A mixed ANOVA on the total scores on the comprehension task was conducted with year group (Year 3 vs. Year 5) as a between subjects factor and presentation format (segmented vs. whole) and question type (local vs. global coherence) as within subjects factors. There was a significant main effect of year group (F (1,58) = 14.03, $p < 0.001$, $\eta_p^2 = .195$): The older children (Year 5 mean = 18.14) obtained significantly higher scores compared with the younger children (Year 3 mean = 16.57). Segmented presentation of the stimulus resulted in slightly higher scores than whole (Ms = 17.77 vs. 16.94), but the effect of presentation format did not reach conventional levels of significance (F(1,58) = 3.76, $p = 0.06$, $\eta_p^2 = .061$). There was a large and significant main effect of question type (F(1,58) = 35.10, $p < 0.001$, $\eta_p^2 = .377$): Scores on the global coherence questions (mean = 18.44) were higher than the local coherence questions (mean = 16.26).

These main effects were qualified by significant two-way and three-way interactions. There was a significant interaction between question type and year group (F(1,58) = 8.91, $p = 0.004$, $\eta_p^2 = .133$). As can be seen from the means in Table 1, there was a bigger increase in scores for the local coherence items with increasing age compared to small increases in scores for global coherence items, and the year groups differed significantly on local coherence inferences (t(58) = 4.594, $p < .001$) but not on global coherence inferences (t(58) = 0.883, $p = .381$). There was a significant interaction between question type and presentation format (F(1,58) = 7.04, $p = 0.010$, $\eta_p^2 = .108$). It can be seen from Table 1 that scores for the local coherence questions were higher for the segmented compared to the whole presentation format (t(59) = 2.760, $p = .008$), but the same was not true for global

coherence inferences ($t(59) = 0.137$, $p = .891$) The interaction between presentation format and year group was not significant ($F(1,58) = 1.49$, $p = 0.23$, $\eta_p^2 = .025$). The two-way interactions were qualified by a significant three-way interaction between question type, presentation format and year group ($F(1,58) = 4.54$, $p = 0.04$, $\eta_p^2 = .073$). This is depicted in Figure 1.

Figure 1 around here

In order to understand the 3-way interaction, separate analyses by year group were conducted. For each year group an ANOVA of scores on the comprehension task was conducted with presentation format (segmented vs. whole) and question type (local vs. global coherence) as within subjects factors. For children in Year 3 there was a significant main of effect of presentation format ($F(1,28) = 4.88$, $p = 0.04$, $\eta_p^2 = .148$). The segmented format (mean = 17.24) resulted in significantly higher scores than whole format (mean = 15.89). There was also a significant main effect of question type (F (1,28) = 36.72, $p < 0.001$, $\eta_p^2 = .567$). Scores on the global coherence questions (mean = 18.21) were higher than the local coherence questions (mean = 14.93). Finally, there was a significant interaction between question type and presentation format (F (1,28) = 15.88, $p < 0.001$, $\eta_p^2 = .362$). Presentation format had a greater effect on the local than global coherence questions, with higher scores for local coherence questions in the segmented compared to whole format. In contrast, scores for global coherence questions were very similar in both presentation formats. In contrast for the children in Year 5, only the main effect of question type was significant (F (1,30) = 4.67, $p = 0.04$, $\eta_p^2 = .135$), because scores on the global coherence questions (mean = 18.67) were higher than scores on the local coherence questions (mean = 17.60). There was no main effect of presentation format (F (1,30) = .27, $p = 0.61$, $\eta_p^2 = .009$)

and no interaction between question type and presentation format (F (1,30) = .11, p = 0.74, η_p^2 = .004).

Discussion

The present study set out to explore listening comprehension skills in typically developing children in whole and segmented format using questions that assessed the ability to establish local and global coherence. We found that age, presentation format, and question type each influenced performance. We discuss each finding in turn and their implications for our understanding of language comprehension and assessment.

Overall, the results show a developmental trend. The older children were better at answering questions that required local and global coherence inferences than the younger children. Therefore this novel approach to assessing listening comprehension is sensitive to developmental progression that would be expected in comprehension. Critically, these age-related differences were not qualified by an interaction with presentation format, indicating that the age-related differences existed independent of presentation format, representing developmental progression in language comprehension in general.

Surprisingly, we did not find a significant benefit overall of one presentation format (whole vs. segmented) over the other. This initially suggests that, in general, breaking the text up does not have a detrimental or an advantageous effect on the ability to build a mental model. However, there was a significant three way interaction between presentation format, age and question type. That interaction arose because the youngest children did most poorly on local coherence inferences in the whole text condition relative to the segmented text; their performance on the global coherence inference questions was higher and was not influenced by presentation format. We discuss possible reasons for this finding below, in our consideration of differences between the two types of question.

To understand the interaction, we conducted analysis by year group and this revealed a different pattern of results for the younger and older children. There was a clear and strong benefit of segmentation for the younger children (Year 3) only. The segmentation may have aided performance by focusing children's attention on coherence because it was overtly 'checked' by the ongoing questions, thereby encouraging a higher standard of coherence (van den Broek et al., 2001b). Alternatively, segmentation may have supported the local coherence questions because it alleviated memory demands because children were prompted to integrate the relevant information close to where it had been presented. The pause between the different segments of the story may have also provided the younger children with an opportunity to monitor coherence so far and attempt to fill in any missing information before presentation of the next segment. Clearly, the advantage for the youngest age group in the segmented presentation condition indicates that younger children are able to make these inferences, even if they do not do so spontaneously during typical listening. We attribute the absence of an effect of presentation format for the older children (Year 5) to their greater experience of comprehension and inherently higher standards of coherence.

In line with previous research (e.g. Currie and Cain, 2015, Lynch et al., 2008, Ackerman, 1988), children were able to resolve both local and global coherence inferences. The results suggest that children can go beyond individual words and sentences and generate local and global coherence inferences to construct a meaning-based representation of the story. Performance overall was stronger on global coherence questions compared to the local coherence questions. The question types are not directly comparable because they assess different elements of the stories. Therefore, we cannot say that children are *better* at establishing global coherence than local coherence, in general,

however we believe that there are clear theoretical reasons for this difference in this particular study, which we expand on below.

The better performance for global coherence inference questions is in line with Currie and Cain's (2015) study (and both studies used some of the same texts). Currie and Cain proposed that children may show stronger performance for global coherence inferences because of key differences in the 'centrality' of the information tapped by the two question types. Some of the global coherence inference questions related to story themes, for example the setting or main character. For that reason, these inferences were essential to understand core details of the text. For the local coherence inferences, the links made between these locally adjacent elements in the text could be viewed as supporting more peripheral details in the text. This explanation is supported by research into the centrality effect, which shows that children are more likely to remember information that is central to the overall meaning of the story than they are to remember peripheral information (Miller and Keenan, 2009).

It is also logical that if children are unable to make all of the required inferences whilst listening to a story, a strategic technique to adopt would be to focus on understanding the overall gist of the story. However, another difference between these questions, which we did not control for in this study, was the number of clues provided in the story to support each inference: local coherence inferences had, by their nature, fewer clues available than did global coherence inferences because these questions required the integration of meanings of specific clauses and sentences within the text. Future work could usefully compare performance on local coherence inferences that are central vs. peripheral to story comprehension and to additionally investigate whether the number or position of clues for each inference type influences performance.

The poorer performance on the local coherence inferences in general was qualified by an interaction with age and presentation format. We consider the interactions in relation to the roles that these inferences may play in comprehension of these stories and the different processing demands involved. The Year 3 children found the local coherence inferences more difficult in the whole presentation format condition than in the segmented format. There are at least two possible reasons for this. First, segmentation may direct attention to the key elements or relationships in the text resulting in superior performance when questions are asked after each relevant segment. As noted above, if these inferences are less central to understanding the gist of the story, comprehenders may have been less likely to generate these spontaneously when listening to the story. For that reason, greater facilitation was found for the local coherence inferences. An alternative, and not mutually exclusive, explanation is that segmentation may enable inference making in a situation with lower memory load than when questions are asked at the end of a text. Both explanations could explain why the younger and less experienced comprehenders were most likely to show this effect. This finding requires further exploration with a broader age range and critically with children with language and communication difficulties to explore fully its implications. Such work should also seek to test between these alternatives to identify precisely why segmentation supports younger children's comprehension. Determination of which explanation best fits our data would be an important step to understanding why younger children and poorer comprehenders fail to generate sufficient inferences to ensure coherence.

There was an interaction between age and question type. The interaction arose because there were small increases in the ability to establish global coherence with increasing age, compared to bigger increases in scores for the local coherence items. We

believe that these age improvements reflect the less central role that the local coherence inferences played in understanding core story details in these materials. The increasing cognitive and strategic resources that come with age may therefore underpin this finding, enabling children to actively make a greater range of inferences during story presentation as they get older. This suggests that in contrast to the suggestion that children process text in a piecemeal manner (Schmidt and Paris, 1983), in fact they may be striving to establish global coherence but resource limitations effect the success of this for younger children. However, as noted earlier, the two types of inference tapped different aspects of knowledge so a direct comparison is not possible.

Limitations

There are a number of limitations of this study that should be addressed in future work. First, whilst we observed developmental changes in comprehension ability, we tested only two age groups. It would be useful to include a broader age range to see how the relationships between the different presentation formats change for more and less competent comprehenders.

Second, we focussed on listening comprehension. Whilst it is well established that reading and listening comprehension draw on many of the same language and cognitive skills (e.g. vocabulary, knowledge of grammar and higher-level language skills) in order to construct a memory-based representation of the text and the Simple View of Reading (Hoover and Gough, 1990) assumes the same processes underlie both reading and listening comprehension, each modality of task also presents different challenges to the language learner. For example listening tasks present additional attentional and memory demands and the stimulus is transient in nature (Cain and Bignell, 2014). Therefore we cannot assume that the results from the present study would be replicated in a reading task. Future

research should compare performance on reading and listening comprehension tasks in the whole and segmented format. It is also important to explore the effect of segmented presentation for children with limited decoding skills (beginning readers) and whether this format is facilitative for comprehension or is burdensome given their limited working memory capacity and poorer decoding skills.

Finally, children did not perform at ceiling levels in either presentation format so the improvements offered by certain conditions (e.g. segmented presentation for Year 3 children) are limited and other cognitive capacities such as memory and attention will have affected performance. It is important for future work to explore the relative influence of these factors as they are likely to have an impact on success in comprehension tasks and may help to inform support strategies for children with impaired comprehension.

Implications for practice

Despite these limitations, the results from this study have a number of potential implications for teachers and other professionals in terms of supporting and assessing comprehension skills. The results suggest that segmented presentation of comprehension tasks results in better comprehension for *some* children. The segmented text presentation format fits well with the UK National Curriculum (Department for Education, 2014) guidelines, which encourage online questioning and discussion to develop comprehension skills. Segmented presentation also has ecological validity because it reflects a child's real life experiences as they will often pause while listening to a story to discuss what is happening with their caregiver/teacher. The success of children's comprehension following adult-guided interaction with text is well documented with very young children (Kang et al., 2009) and shared book reading is hypothesised to develop children's use and comprehension of complex constructions (Cameron-Faulkner and Noble, 2013) and to make

causal elaborations (Makdissi and Bosclair, 2006), which would be beneficial to poor comprehenders. The segmented presentation format in this study could be viewed as a more 'structured' version of shared book reading as set questions are asked after each segment of text and we advocate the utility of this approach for both assessing a child's inference making potential and supporting their text comprehension.

We note that standardised assessments of reading and listening comprehension typically follow a whole text presentation format and do not enable clinicians and teachers to assess the *potential* of inference making ability during text presentation, even though this format is used informally to support the construction of meaning from text, as noted above. In addition, standardised assessments do not systematically distinguish between local and global coherence inferences, even though these have been distinguished empirically (Cain and Oakhill, 2014; Currie and Cain, 2015) and may have different developmental trajectories, as suggested by our data. Our findings suggest that these factors are useful to identification of specific difficulties and should be considered in test development in the future.

In sum, the present study explored the use of a novel approach to comprehension assessment including assessment of ability to make both global and local inferences in segmented vs. whole presentation format. The results showed that segmented presentation may be beneficial for some children, particularly for the generation of inferences to support local coherence. Further studies are needed to confirm this effect and explore whether it remains in different modalities, across a wider age range and most importantly in children with comprehension or language and communication difficulties. However the results from the present study have potential implications for the design of comprehension assessments

that can measure *potential* in addition to actual performance as well as for classroom practice supporting children with comprehension difficulties in school.

Acknowledgements

We thank all of the teachers and pupils at the participating schools for their help with this research; and Belinda Buckley, Gemma Curtis, Susan Dowd, Eva Kristanova, Susan Massey, Kirsty Meehan and Lucy McFarlane for their help with data collection, developing the scoring rubric, and scoring and data inputting.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

ACKERMAN, B. P. 1986. Referential and causal coherence in the story comprehension of children and adults. *Journal of Experimental Child Psychology,* 41**,** 336-366.

ACKERMAN, B. P. 1988. Reason inferences in the story comprehension of children and adults. *Child Development,* 59**,** 1426-1442.

BAKER, L. & STEIN, N. 1981. The development of prose comprehension skills. *In:* SANTA, C. M. & HAYES, B. L. (eds.) *Children's prose comprehension: Research and practice.* Newark, DE: International Reading Association.

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.

BISHOP, D. V. & ADAMS, C. 1992. Comprehension problems in children with specific language impairment: literal and inferential meaning. *Journal of Speech & Hearing Research,* 35**,** 119-29.

CAIN, K. 2009. Making sense of text: skills that support text comprehension and its development. *Perspectives on Language and Literacy,* 35**,** 11-14.

CAIN, K. & BIGNELL, S. 2014. Reading and listening comprehension and their relation to inattention and hyperactivity. *British Journal of Educational Psychology,* 84**,** 108-124.

CAIN, K. & OAKHILL, J. 1999. Inference making ability and its relation to comprehension failure in young children. *Reading and Writing,* 11**,** 489-503.

CAIN, K. & OAKHILL, J. 2014. Reading comprehension and vocabulary: Is vocabulary more important for some aspects of comprehension? *L'Année Psychologique,* 114**,** 647-662.

CAIN, K., OAKHILL, J. & BRYANT, P. 2000. Investigating the causes of reading comprehension failure: The comprehension-age match design. *Reading and Writing,* 12**,** 31-40.

CAIN, K., OAKHILL, J. V., BARNES, M. A. & BRYANT, P. E. 2001. Comprehension skill, inference making ability and their relation to knowledge. *Memory and Cognition,* 29**,** 850-859.

CAMERON-FAULKNER, T. & NOBLE, C. 2013. A comparison of book text and Child Directed Speech. *First Language,* 33**,** 268-279.

CAMILLERI, B. & LAW, J. 2007. Assessing children referred to speech and language therapy: Static and dynamic assessment of receptive vocabulary. *International Journal of Speech-Language Pathology,* 9**,** 312-322.

CARLSON, S. E., VAN DEN BROEK, P., MCMASTER, K., RAPP, D. N., BOHN-GETTLER, C. M., KENDEOU, P. & WHITE, M. J. 2014. Effects of Comprehension Skill on Inference Generation during Reading. *International Journal of Disability Development and Education,* 61**,** 258-274.

CURRIE, N. K. & CAIN, K. 2015. Children's inference generation: the role of vocabulary and working memory. *Journal of Experimental Child Psychology,* 137**,** 57-75.

DEPARTMENT FOR EDUCATION. 2014. *Statutory guidance. National curriculum in England: English programmes of study* [Online]. Available: www.gov.uk/government/publications/national-curriculum-in-england-english-programmes-of-study/national-curriculum-in-england-english-programmes-of-study [Accessed 11th November 2014].

GARNHAM, A. 2010. Models of processing: discourse. *Wiley Interdisciplinary Reviews: Cognitive Science,* 1**,** 845-853.

GRAESSER, A. C., SINGER, M. & TRABASSO, T. 1994. Constructing inferences during narrative text comprehension. *Psychological Review,* 101**,** 371-395.

HASSON, N., DODD, B. & BOTTING, N. 2012. Dynamic Assessment of Sentence Structure (DASS): design and evaluation of a novel procedure for the assessment of syntax in children with language impairments. *International Journal of Language & Communication Disorders,* 47**,** 285-299.

HASSON, N. & JOFFE, V. 2007. The case for dynamic assessment in speech and language therapy. *Child Language Teaching and Therapy,* 23**,** 9-25.

HOOVER, W. A. & GOUGH, P. B. 1990. The simple view of reading. *Reading and Writing,* 2**,** 127-160.
JOHNSON-LAIRD, P. N. 1980. Mental models in cognitive science. *Cognitive Science,* 4**,** 71-115.
KANG, J., KIM, Y. & PAN, B. 2009. Five-year-olds' book talk and story retelling: Contributions of mother-child joint bookreading. . *First Language,* 29**,** 243-265.
KENDEOU, P. 2014. Validation and Comprehension: An Integrated Overview. *Discourse Processes,* 51**,** 189-200.
KENDEOU, P., VAN DEN BROEK, P., HELDER, A. & KARLSSON, J. 2014. A cognitive view of reading comprehension: Implications for reading difficulties. *Learning Disabilities Research and Practice,* 29**,** 10-16.
KINTSCH, W. 1998. *Comprehension: A Paradigm for Cognition,* New York, NY, Cambridge University Press.
LAING, S. P. & KAMHI, A. G. 2002. The use of think-aloud protocols to compare inferencing abilities in average and below-average readers. *Journal of Learning Disabilities,* 35**,** 436-447.
LARRC 2015. The dimensionality of language ability in young children. *Child Development,* 86**,** 1948-1965.
LONG, D. L. & CHONG, J. L. 2001. Comprehension skill and global coherence: A paradoxical picture of poor comprehenders' abilities. *Journal of Experimental Psychology-Learning Memory and Cognition,* 27**,** 1424-1429.
LONG, D. L., OPPY, B. J. & SEELY, M. R. 1994. Individual-differences in the time-course of inferential processing. *Journal of Experimental Psychology-Learning Memory and Cognition,* 20**,** 1456-1470.
LYNCH, J. S., VAN DEN BROEK, P., KREMER, K., KENDEOU, P., WHITE, M. J. & LORCH, E. P. 2008. The development of narrative comprehension and its relation to other early reading skills. *Reading Psychology,* 29**,** 327-365.
MAKDISSI, H., & BOSCLAIR A. (2006). Interactive reading: A context for expanding the expression of causal relations in preschoolers. *Written Language and Literacy, 9*, 177-211. doi: 10.1075/wll.9.2.02mak
MILLER, A. C. & KEENAN, J. M. 2009. How word decoding skill impacts text memory: The centrality deficit and how domain knowledge can compensate. *Annals of Dyslexia,* 59**,** 99-113.
NATION, K., CLARKE, P., MARSHALL, C. M. & DURAND, M. 2004. Hidden language impairments in children: Parallels between poor reading comprehension and specific language impairment? *Journal of Speech Language and Hearing Research,* 47**,** 199-211.
NEALE, M. D. 1997. *Neale Analysis of Reading Ability II: Second Revised British Edition*, Windsor, England: NFER-Nelson.
OAKHILL, J. 1984. Inferential and memory skills in children' comprehension of stories. *British Journal of Educational Psychology,* 54**,** 31-39.
OLSON, G. M., DUFFY, S. A. & MACK, R. L. 1985. Question-asking as a component of text comprehension. *In:* GRAESSER, A. C. & BLACK, J. B. (eds.) *The psychology of questions.* Mahwah, NJ: Erlbaum.
SCHMIDT, C. R. & PARIS, S. G. 1983. Childrens use of successive clues to generate and monitor inferences. *Child Development,* 54**,** 742-759.
SEMEL, E., WIIG, E. H. & SECORD, W. 2006. *Clinical Evaluation of Language Fundamentals - Fourth Edition UK (CELF-4 UK),* London, Harcourt Assessment
SNOWLING, M. J., STOTHARD, S., CLARKE, P., BOWYER-CRANE, C., HARRINGTON, A., TRUELOVE, E., NATION, K. & HULME, C. 2009. *York Assessment of Reading for Comprehension (YARC)*, GL assessment.
VAN DEN BROEK, P., LORCH, R. F., LINDERHOLM, T. & GUSTAFSON, M. 2001a. The effects of readers' goals on inference generation and memory for texts. *Memory & Cognition,* 29**,** 1081-1087.
VAN DEN BROEK, P., RISDEN, K. & HUSEBYE-HARTMAN, E. 1995. The role of readers' standards for coherence in the generation of inferences during reading. *In:* LORCH, R. F. & O'BRIEN, E. J. (eds.) *Sources of coherence in reading.* Hillsdale, NJ: Lawrence Erlbaum.

VAN DEN BROEK, P., TZENG, Y., RISDEN, K., TRABASSO, T. & BASCHE, P. 2001b. Inferential questioning: Effects on comprehension of narrative texts as a function of grade and timing. *Journal of Educational Psychology,* 93**,** 521-529.

WECHSLER, D. 2005. *Wechsler Individual Achievement Test. Second UK Edition,* London, Harcourt Assessment.

Appendix A. Example story with comprehension questions. This shows the structure of the story presentation and questions in the segmented format (for the whole format, the whole story is presented, followed by each of the questions in turn). (G) refers to a question tapping a global inference and (L) refers to a question tapping a local inference.

Birthday

Today was Grandma's birthday. The family was getting ready for the party. Dad and Josh were putting up the party tent in the back garden. Mum told them to put on some suncream, so that they didn't burn.

Q1. What were the family getting ready for? (L)

Q2. What was the weather like? (G)

Mum drove over to pick up Grandma, who lived an hour away. Mum told Sarah to keep an eye on the cake in the oven and to make some fruit juice. Sarah was slicing oranges when the knife slipped. Her finger was bleeding but she couldn't find any plasters! Luckily, Brenda, their next-door neighbour, had some.

Q3. Why did Sarah need some plasters?(G)

Q4. Where did Sarah get the plasters? (L)

Back in the house, the kitchen was filled with smoke. Sarah looked in the oven. Oh dear! Mum would be mad.

Q5. Why would Mum be mad? (G)

Then, Sarah had an idea. She drove to the supermarket. When she got back home, her

aunts, uncles, and cousins were all waiting quietly in the party tent. Sarah put what she had bought at the centre of the dessert table.

Q6. Why did Sarah drive to the supermarket? (L)

Q7. Where was the dessert table? (L)

A few minutes later, Mum walked into the party tent with Grandma. Everything looked perfect. Grandma was amazed that all of her family was there. "What a wonderful surprise." she said.

Q8. Why was Grandma surprised?(G)

Appendix B. Scoring rubric with example answers that would score 2 points (correct response), 1 point (partially correct) and 0 points (incorrect) for Birthday story.

	Question	**Correct (2 points)**	**Partially correct (1 point)**	**Incorrect (0 points)**
Q1	What were the family getting ready for?	Grandma's (birthday) party	A party	A holiday
Q2	What was the weather like?	Sunny	Hot	Rainy
Q3	Why did Sarah need some plasters?	cut her finger (on knife)	finger was bleeding / knife slipped and she cut herself	
Q4	Where did Sarah get the plasters?	next door / from Brenda / from neighbour	from a friend's house	From a shop
Q5	Why would Mum be mad?	cake was burnt or ruined / Sarah burnt the cake	she burnt the cake/ something burnt	because the kitchen was full of smoke
Q6	Why did Sarah drive to the supermarket?	to get/buy a (new) cake/she needed a new cake	because the cake was burnt / to get food for the party	because she needed something to put on the cake / to get some fruit
Q7	Where was the dessert table?	in (party) tent	inside	in the kitchen /near the party tent
Q8	Why was Grandma surprised?	didn't know they were having a party for her/ no one told her there was a surprise party	all the family was there/ it was a surprise party	it looked neat and tidy / because everybody got dressed up

Table 1. Mean scores (standard deviation) out of 24 on whole and segmented versions of the comprehension task, split into scores for the local and global questions.

	Year 3 (n=29)		**Year 5 (n=31)**	
	Local	**Global**	**Local**	**Global**
Whole	13.45 (3.17)	18.34 (2.64)	17.35 (3.29)	18.61 (3.04)
Segmented	16.41 (3.04)	18.07 (2.74)	17.84 (3.33)	18.74 (2.77)

Figure 1: Graph depicting the three-way interaction between year group, presentation format and question type (mean sum score and standard error bars).