Assessment of inference making in children using comprehension questions and story retelling: Effect of text modality and story presentation format

Jenny Freed*

University of Manchester

Kate Cain

Lancaster University

Dr Jenny Freed: Ellen Wilkinson Building, Human Communication, Development and Hearing, 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.

Running Head: Assessment of inference making in children

Abstract

Background. Reading and listening comprehension are essential for accessing the school curriculum. Inference making is integral to successful comprehension and involves integrating information between clauses (local coherence) and integrating information with background knowledge (global coherence). We require appropriate methods to assess comprehension and inference making in order to identify areas of difficulty and provide appropriate support.

Aims. Typically developing children's ability to generate local and global coherence inferences was assessed. The effect of text modality (reading and listening comprehension) and presentation format (stories presented in segmented and whole story format) was explored using two comprehension measures (question answering and story retell). The main aims were to determine whether there were advantages for reading or listening comprehension and for segmented or whole text presentation.

Methods & Procedures. Typically developing children in Year 3 (n=33) and Year 5 (n=40) either read or listened to short stories. Their ability to generate global and local coherence inferences was assessed in two ways: answers to inference-tapping questions and story retelling (scored for inclusion of necessary inferences). Stories were presented either in whole format (all questions after the story) or segmented format (questions asked at specific points during story presentation); the retelling was always after the complete story and questions had been presented.

Outcomes & Results. For both comprehension measures, there was developmental progression between age groups and a benefit for the reading modality. Scores were higher for global coherence than local coherence inferences, but the effect was significant only for

3

the question answering responses, not retells. For retells there was a benefit in presenting the text as a whole compared with the segmented format, but this effect was not present for the comprehension questions. There was a significant interaction between inference type and modality for both comprehension measures (question answering and story retell): for the local coherence inferences scores were significantly greater in the reading compared to the listening modality, but performance on the global coherence inferences did not differ significantly between modalities.

Conclusions & Implications. Clinicians, teachers and other professionals should consider the modality and presentation format for comprehension tasks to utilise areas of strength and support areas of difficulty. Oral presentation may result in poorer comprehension relative to written presentation in general, and may particularly affect local integrative processing. These findings have important implications for the development of appropriate assessments as well as for supporting children with comprehension difficulties.

Keywords: Inference, Local coherence, Global coherence, School-aged children, Reading, Listening, Story retell

What is already known on this subject

Reading and listening comprehension are critical for accessing the school curriculum and educational success. Inference making is integral to successful comprehension and involves integrating information between clauses (local coherence) as well as integrating information with background knowledge (global coherence). Children have an awareness of the need to generate coherence inferences but not all children will generate sufficient coherence inferences for adequate comprehension during text presentation. Existing assessment tools measure comprehension by asking questions after story presentation. This provides an

overall indication of comprehension or inference making ability and can identify children with comprehension or language and communication difficulties.

What this study adds

The study compared coherence inference making in two presentation conditions: whole format (all questions after the story) or segmented format (questions asked at specific points during story presentation). Children (aged seven to ten years) were assessed in the reading or listening modality. Two comprehension measures were used (inference-tapping questions and story retell). There was developmental progression and a benefit for the reading modality for both comprehension measures. Scores were higher for global coherence than local coherence inferences for the comprehension questions. There was a benefit in presenting the text as a whole compared with the segmented format for story retells.

Clinical implications of this study

The results are important for clinicians and other professionals assessing and supporting comprehension skills. The results suggest that the modality and presentation format of comprehension tasks should be considered to utilise areas of strength and support areas of difficulty. The optimum form of input and structure may depend on a child's individual profile and the skill being assessed or supported. Targeted questions may identify a child's potential to generate an inference. This may assist identification of children who may require more targeted or specialist intervention. The reading modality may provide a means of support for development of verbal comprehension.

Assessment of inference making in children using comprehension questions and story

retelling: Effect of text modality and story presentation format

Reading and listening comprehension are critical for classroom learning and educational

success (DCSF, 2008). In the classroom, children are required to understand texts that they

read themselves (reading comprehension), as well as those read aloud to them (listening

comprehension). Therefore, successful discourse comprehension in both modalities is

important to access the curriculum. As a result, reading and listening comprehension are

prioritised in the National Curriculum for England (Department for Education, 2014)

programmes of study for Key Stage 1 (age 5-7) and Key Stage 2 (age 7-11). There is

considerable variation in reading comprehension skill in developing readers, and the

prevalence of reading comprehension difficulties have been described as 'relatively

common' (Hulme & Snowling, 2011). Recently, between 7-10% of children in this age range

have been identified with poor comprehension skills despite having adequate word reading

(decoding) or nonword reading skills (e.g., Elwér et al., 2015) and the percentage identified

may depend on the selection criteria (Keenan et al., 2014) and also age of participants (Catts

et al., 2012). Therefore, it is crucial to gain a better understanding of the factors that

influence listening and reading comprehension to support comprehension development and

intervention. In this study we add to the literature by examining children's ability to make

coherence inferences to support comprehension of both spoken and written text, and

whether presentation format influences performance. This work has important implications

for teachers, speech and language therapists and other professionals assessing and

supporting comprehension skills.

Coherence inferences and standards of coherence

Coherence is central to successful discourse comprehension. When readers and listeners go beyond individual words and simple sentences, they engage in integrative and constructive processing to build a coherent representation of the text's meaning: a mental model or situation model (Kintsch, 1998). In this study, we examined children's ability to generate two types of inference that are necessary to establish coherence of the mental model: local and global coherence inferences (Graesser, Singer, & Trabasso, 1994). In order to generate a local coherence inference, readers and listeners use the information presented in the text, by combining information from successive clauses and sentences within the text. These inferences may be cued by pronouns, or can be made by appreciating the link between synonyms or category exemplars (Cain & Oakhill, 1999), for example: inferring that the drink was orange juice from: 'He fetched a glass of orange juice. The drink was very refreshing.' Global coherence inferences, in contrast, require integration of information in the text with the child's background knowledge (Cain & Oakhill, 1999). This may involve inferring motivations or establishing a theme or character identity within a text, for example inferring a school setting in a story from related words: teachers, classroom, lessons. Children with poor reading and language comprehension have problems with both types of inference (Bishop & Adams, 1992; Cain & Oakhill, 1999). Children's ability to generate these two types of inferences are related, but also theoretically and statistically separable (Language and Reading Research Consortium & Muijselaar, 2018).

Language comprehension is a dynamic process and coherence inferences will be made in order to integrate discourse information into the mental model of the text constructed so far. A failure to generate a necessary inference may arise if the comprehender has a weak standard of coherence (van den Broek, Lorch, Linderholm, & Gustafson, 2001; van den

Broek, Risden, & Husebye-Hartman, 1995). If they process text on a word-by-word or sentence-by-sentence level rather than attend to discourse-level meaning (Schmidt & Paris, 1983) they will adopt a different standard of coherence for comprehension to older or more skilled comprehenders (van den Broek et al., 1995). Therefore, although children in the age range studied here (7 to 10 years old) have an awareness of the need to generate coherence inferences and are able to do so when prompted (Ackerman, 1986; Lynch et al., 2008), not all children will generate sufficient coherence inferences for adequate comprehension during text presentation. A greater understanding of developmental change in the generation of necessary coherences inferences is needed to inform both theory and classroom practice.

Presentation format

Existing assessment tools that measure comprehension after story presentation provide an overall indication of comprehension or inference making ability and can be useful for identifying children with comprehension or language and communication difficulties (Bishop & Adams, 1992; Cain, Oakhill, & Bryant, 2000). However, such assessments cannot identify where, in the course of the story, failures to generate necessary inferences occur. An innovation in this study was to compare coherence inference making in two presentation conditions: in the segmented presentation condition, questions that prompted inference making were asked at specific points during story presentation immediately after the relevant information had been presented; in the whole format presentation condition, questions were asked at the end of each story. Asking questions at the end of the text is typical in developmental studies of inference making (e.g. Bishop & Adams, 1992; Cain & Oakhill, 1999), which limits our ability to identify the locus of inference making difficulties.

The segmented text technique provides a window into the comprehension process as the text is presented and the representation of the text's meaning is constructed, incrementally (Carlson et al., 2014; Freed & Cain, 2017). Of note, this approach reflects National Curriculum guidelines where 'online' questioning and discussion during reading instruction are utilised to develop comprehension skills (Department for Education, 2014). This technique also overlaps with a dynamic approach to assessment often used by speech and language therapists assessing children who have language disorders in schools (Camilleri & Law, 2007; Hasson, Dodd, & Botting, 2012; Hasson & Joffe, 2007). A dynamic approach is a facilitative technique in which the speech and language therapist uses prompts to scaffold and support the child and identify their potential on a specific language skill.

Carlson et al. (2014) compared inference generation *during* story presentation (segmented approach) between 8-10 year old children with different levels of comprehension skill. Participants read stories aloud and were asked questions at specific points in the story. They found that poor readers made a similar number of inferences to good readers suggesting that poor readers are capable of making inferences but their apparent difficulties with discourse comprehension arise because they fail to make inferences when needed as text is presented. However, Carlson et al did not compare performance during text presentation with performance after text presentation, so their results do not indicate whether or not there would be a difference between the two presentation formats for each group. Freed and Cain (2017) compared typically developing children's listening comprehension skills for inference generation during story presentation (segmented approach) with inference generation after story presentation (whole text approach). Seven- to ten-year-old children listened to stories in both presentation formats

and answered questions that required either global or local coherence inferences. A benefit was apparent for the segmented presentation format, particularly for younger children (7-8 years) and their ability to generate local coherence inferences; there was no apparent benefit for global coherence inferences.

In both studies, the segmented approach may have scaffolded comprehension by focusing attention on certain points in the story prompting links to be made between elements of the text and an inference to be made (Freed & Cain, 2017; van den Broek, Tzeng Risden, Trabasso, & Basche, 2001). Therefore, if a child's difficulties in traditional format tasks relate to a failure to make inferences *as needed* during text presentation, better performance should be apparent when texts are presented in segments, with inference questions asked at the specific point in the text where required. Children may adapt their standards of coherence according to the task demands (van den Broek, Lorch, et al., 2001; van den Broek et al., 1995). Therefore the segmented format may reveal a child's *potential* to generate inferences. The segmented format may also increase inference generation by reducing the memory load – the amount of information to be remembered (Olson, Duffy, & Mack, 1985; van den Broek, Tzeng et al., 2001).

Modality

Reading and listening comprehension draw on many of the same language and cognitive skills in order to construct a mental representation of the text, such as vocabulary, knowledge of grammar and higher-level language skills such as inference making (Currie & Cain, 2015; Kendeou et al., 2009; Lepola et al., 2012). Influential models of reading, such as the Simple View of Reading (Hoover & Gough, 1990), assume that the same language processes underlie comprehension in both modalities. However reading and listening

comprehension tasks make different demands on memory, attention, and processing skills (Cain & Bignell, 2014; Language and Reading Research Consortium, Jiang, & Farquharson, 2018). Text spoken aloud is transient so accurate moment-by-moment processing and sustained attention to the speaker is essential, because the text cannot be reviewed as it can be in print. This is reflected in the fact that attention and memory are stronger predictors of listening than reading comprehension (Cain & Bignell, 2014; Language and Reading Research Consortium et al., 2018). Another potential disadvantage of spoken text is that the speaker, not the listener, controls the pace of delivery. When children are allowed to pace their own reading, rather than listen to speech, comprehension is better, perhaps because children with slow language processing are not at such a disadvantage (Casteel, 1993). Due to the differences in the memory, attention, and processing demands of reading and listening comprehension, we compared the two and sought to determine if segmentation had the same effect in each modality. There are no studies, to our knowledge that have compared the two presentation formats in different modalities. We used different groups of children for each modality of task. This is the most appropriate way to compare performance in the two modalities to avoid practice effects. This also meant that three stories could be completed in each presentation format allowing within subject comparisons between whole and segmented presentation and local and global coherence inferences.

Measurement

Comprehension of text is typically assessed by questions, as noted above. In this study, children additionally were asked to retell each story after the (final set of) comprehension questions. Retells can reveal what has been encoded in the mental model that the child has constructed (Cataldo & Oakhill, 2000). One might hypothesise that if encouraged to actively

engage with the story through segmented presentation, rather than being questioned after the whole text, the child may produce richer and more complex reconstructions of the story, demonstrating a more complete situation model. In young children (aged 4-6) Silva and Cain (2019) reported that narratives produced after comprehension questions were more coherent than those produced before questions using a picture sequence task. However, van den Broek, Tzeng, et al. (2001) reported that questioning during text reading resulted in poorer story retells compared to story retells with no questioning for 9-10 year old children. They suggested that the disjointed nature of the segmented questioning might interfere with the child's overall construction of a mental model of the stimulus material and affect their limited attentional resources. Thus, studies to date with different age groups present contradictory findings of the influence of segmented questioning; we know of no studies that have compared the same materials and procedure between two different age groups. By including two different measures of comprehension, and contrasting two different age groups, we were able to explore how modality and presentation format influenced inference making when prompted by questions, and the encoding of this information into the mental model of the text.

Aims

We investigated comprehension and story retelling skills in typically developing children between seven and ten years old using an assessment that compared whole vs segmented presentation of passages and contrasted reading vs listening task modality. Comprehension questions to assess target global and local coherence inferences and analysis of story retells for these inferences were used to assess the extent to which children could generate and encode coherence inferences for a narrative text.

Our broad aims were: (1) to determine whether segmented presentation benefited inference making and to determine whether the effects were the same for reading and listening versions of the task, and (2) to establish whether the effects were consistent across the primary school age range. We analysed both the ability to answer the inference-tapping questions and also the retells for the inclusion of inferences, to assess both the ability to generate inferences and also to encode these into the mental model of the text. This study builds on Freed and Cain (2017) with the addition of the reading modality and the story retell task. We also aimed (3) to determine if the benefits of segmented text can be reproduced with a different sample in the listening modality, and whether the pattern of results is the same for the reading modality.

Method

Ethical approval for this study was granted by The University of Manchester Research Ethics Committee. Consent was obtained from the Headteacher of participating schools. Parents gave informed consent on behalf of their children and children gave assent at the start of each assessment session.

Participants

The participants were 73 children (39 boys) from mainstream primary schools in the Greater Manchester area: 33 were in Year 3 classrooms and were aged 7 to 8 years (18 boys; M =8 years, 4 months, SD= 3.50 months, range = 7;10 – 8;9); and 40 were in Year 5 classrooms and aged 9 to 10 years (21 boys; M = 10 years, 3 months, SD = 3.31 months, range = 9;9 – 10;9). Children who were unable to speak or understand English were excluded as well as any children with known learning, speech or language disorders 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). Children were randomly assigned to complete either the listening (Year 3: n = 17, Year 5: n = 20) or reading (Year 3: n = 16, Year 5: n = 20) comprehension task.

Measures

Children completed assessments across two sessions: two assessments of the comprehension task, one in segmented format and one in whole format; an assessment of receptive vocabulary; and an assessment of word reading accuracy (completed by participants doing the reading modality of the comprehension task only). The materials and procedure for each are described in turn.

Comprehension task. The comprehension task consisted of six short stories similar to those used in school-based comprehension tasks (Department for Education, 2014). The stories used were kindly made available by the Language and Reading Research Consortium (Language and Reading Research Consortium, 2015a) and were also used by Freed and Cain (2017). Each story had eight comprehension questions, four assessing the ability to generate local coherence inferences (testing use of information provided by integration of clauses and sentences within the text) and four assessing global coherence inferences (requiring the integration of information in the text with the child's background knowledge). The order of the questions followed the order of information presented in the story. Freed and Cain (2017) adapted the stories for their study by developing segmented versions of the stories and making minor modifications to vocabulary (further information can be found in Freed and Cain (2017)) and these adapted stimuli were used in this study.

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). In the whole presentation procedure, the child listened to/read each story in turn. At the end of each story they answered eight questions assessing the ability to generate local and global coherence inferences. In the segmented presentation procedure, the child listened to/read each story in five segments. After each segment of the story the child was asked one or two comprehension questions (eight in total for the story). Therefore questions were either presented at the end of the story (whole version) or at intervals during the story (segmented version).

The six stories were 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 (whole/segmented) and the order that they were completed in (whole/segmented, A/B) was counterbalanced independently (see Appendix A for example story). Different groups of children were presented with the stories in the reading format and listening format. This design feature minimised practice effects and also the time spent out of the classroom. For children completing the listening modality, stories were pre-recorded and presented using a PowerPoint file on a laptop computer. This ensured that all children in this condition received the same input (i.e. to control for pace of delivery, word stress etc.). Children who completed the task in the reading modality were presented with the story on a laptop computer via PowerPoint. In the segmented version the text appeared one section at a time. Earlier sections of text remained on the screen when subsequent sections were presented as well as during the comprehension questions. The text was removed prior to the story recall. Children read aloud at their own pace. The number of word reading errors was noted. The average number of words per story was 157.

Children were not presented with any pictures so they could only use information in the text and their own knowledge to generate the coherence inferences.

Comprehension questions. In both modalities, 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. In line with other assessments, children were not told whether they had answered questions correctly, but positive encouragement was given throughout. Responses were written down verbatim for later scoring.

Story retells. After all the questions were asked for each story, children were asked to retell the story imagining they were telling it to someone who has not seen/heard the story. During retelling the researcher offered general encouragement and used general prompts (e.g. "What happened next?"). When the child finished telling the story, the researcher said "That was a good story". Story retells were audio recorded for later transcription and scoring.

Vocabulary. The British Picture Vocabulary Scale: Second Edition (BPVS-2; Dunn, Dunn, Whetton, & Burley, 1997) was used to indicate vocabulary skills in each group. In the BPVS-2 the child hears series of words of increasingly later acquisition and, for each, is asked to choose a corresponding picture from a selection of four. The administration and scoring procedure outlined in the manual was followed.

Word reading accuracy. Word reading accuracy was assessed using the basic reading subtest from the Wechsler Objective Reading Dimensions (WORD; Rust, Golombok, & Trickey, 1993). This subtest requires the child to read out loud a series of individual words presented out of context. This measure was only completed by children who were doing the

reading modality of the comprehension task and was used to ensure that the child had an adequate level of reading ability to access the written materials.

Procedure

Children were seen individually by the researcher in a quiet room on two separate occasions (a minimum of two days apart). In each session the participant was presented with one set of comprehension stories (Set A/Set B). Each set consisted of one practice story to familiarise the child with the task, followed by three test stories. In one of the sessions the stories were presented in segmented format and in the other they were presented in whole format. Thus, all children completed both presentation formats. The order of the sessions was counterbalanced. In the first assessment session, all children also completed the BPVS-II and those doing the reading modality also completed the WORD subtest. The second session consisted of the comprehension task only, in whole or segmented form, depending on the format completed in the first session.

Scoring

Comprehension questions. 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). Each response was coded as correct (2 points), partially correct (1 point) or incorrect (0 points) resulting in a maximum possible score of 8 points each for local and global coherence inferences for each story (4 questions x 2 points for each), and 24 points in total for each set of stories. 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. There were disagreements on scoring for only 17 out of 3504 responses (73 children x 6 stories x 8 questions), showing an agreement level of 99.51%. The 17 disagreements related to 12/48 individual questions from 3/6 stories. The

Weighted Kappa for scoring agreement between the two raters was excellent (=0.99; Landis & Koch, 1977). Final scores were agreed between the two researchers through discussion in order to provide a final set of scores for analysis.

Story retell. Each story retell was audio recorded and later transcribed verbatim. A scoring rubric was developed which provided examples of fully stated, partially stated and not stated/incorrect narrative for each inference. Retells were scored according to whether children included information in their retell that demonstrated they had made the inferences related to each of the comprehension questions, using the same range of scores (correct = 2 points, partially correct = 1 point or incorrect = 0 points) as for the comprehension questions. For example, in the example story (Appendix A), the final question is 'why was Grandma surprised?'. Children are expected to infer that Grandma was surprised because she did not know she has having a party. Children would score two points if their retell included the fact that Grandma was surprised as well as something relating to her mental state of not knowing about the party, for example 'Grandma was surprised because she didn't know they were having a party for her'. Children scored one point if their retell did not include information about Grandma not knowing or expecting the party, for example 'Grandma was surprised because of the party'. If the child's retell only included a verbatim repetition of the text in the story ('Grandma was amazed that all of her family was there. What a wonderful surprise she said') it was scored zero as the child had not demonstrated they had explicitly made the inference. Children would also score zero for incorrect information, for example saying she was surprised because it looked tidy outside. Ten percent of retell data was scored by a second researcher blind to the aims and hypotheses of the study to assess interrater reliability. The Weighted Kappa for scoring agreement between the two raters was excellent (=0.84; Landis & Koch, 1977). Scores from

the first rater were used in the analysis as they had scored retells for all participants. Full

details about scoring are available from the first author on request.

<div align="center">**Results**</div>

We first present descriptive statistics on our background reading and language measures

and then present analyses that address our research aims.

Descriptive statistics. Different children completed the listening and reading tasks, so we

first checked that the groups did not differ significantly on critical variables (see Table 1 for

means and standard deviations). T-tests confirmed that there were no significant

differences between modality groups for chronological age (t(71) = -.04, p = 0.965) or

receptive vocabulary, as measured by standardised score on the BPVS-II (t(71) = -.76, p =

0.447). There was also no significant difference between children in Years 3 and 5 in terms

of standardised score on the BPVS-II (t(71) = -.54, p = 0.588) or standardised score on the

WORD (t(24.43)[1] = .27, p = 0.788). As is clear from the Table 1 means, performance on the

BPVS-II and WORD was in line with population norms. As expected, the younger children

made significantly more word reading errors than did the older children on the

experimental reading comprehension measure (t(34) = 2.16, p = 0.038).

<div align="center">Table 1 around here</div>

Performance on comprehension questions. Table 2 shows the mean scores (and standard

deviations) on the comprehension task split according to modality, presentation format,

year group and inference type. A mixed ANOVA of total scores was conducted with year

group (year 3 vs year 5) and modality (listening vs reading) as between subjects factors and

---

[1] Levene's test due to unequal variances.

presentation format (segmented vs whole) and inference type (local vs global coherence) as within subjects factors.

There was a significant main effect of year group ($F(1,69) = 28.91$, $p < 0.001$, η^2_p = .295), because the older children (Year 5 mean = 19.53) obtained significantly higher scores compared with the younger children (Year 3 mean = 16.84). There was also a significant main effect of modality ($F(1,69) = 14.09$, $p < 0.001$, η^2_p = .170), because scores for children completing the reading modality (mean = 19.12) were significantly higher than for those completing the listening modality (mean = 17.24). The main effect of presentation format was not significant ($F(1,69) = 1.17$, $p = 0.28$, η^2_p = .017): Mean scores for the whole and segmented format were 18.00 and 18.37 respectively. There was a significant main effect of inference type ($F(1,69) = 7.48$, $p = 0.008$, η^2_p = .098): children obtained higher scores for the global coherence (mean = 18.59) than local coherence questions (mean = 17.77). Only one interaction reached significance: inference type and modality ($F(1,69) = 15.95$, $p < 0.001$, η^2_p = .188). All other interactions were not significant: all Fs < 3.90, ps > .05.[2]

The significant interaction between inference type and modality is depicted in Figure 1. Post hoc t-tests indicated that the interaction arose because performance on the global coherence inferences did not differ between modalities ($t(71) = -1.24$, $p = 0.220$), but did differ for local coherence inferences ($t(71) = -4.26$, $p < 0.001$; threshold for multiple comparisons = 0.0125), for which scores were significantly greater for the reading (mean = 38.92) compared to the listening modality (mean = 32.78).

Figure 1 around here

---

[2] Given that decoding (word reading accuracy) is a key skill required for reading comprehension we re-ran the analysis excluding children (two in Year 3 and two in Year 5) who had poor word reading accuracy ability (defined as a score at least one standard deviation below the population norms (standard score <85)). The pattern of results was the same for this smaller group so we report results from the whole group analysis.

Table 2 around here

Performance on story retells. Table 3 shows the mean scores (and standard deviations) for story retells split according to modality, presentation format, year group and inference type. Story retell data was missing for two children in Year 5 (one in each modality) due to equipment failure. A mixed ANOVA of total scores was conducted with the same design as above: year group (year 3 vs year 5) and modality (listening vs reading) as between subjects factors and presentation format (segmented vs whole) and inference type (local vs global coherence) as within subjects factors.

Similar to the analysis of question answering data, there was a significant main effect of year group ($F(1,67) = 7.49$, $p = 0.008$, $\eta^2_p = .101$), because the older children (Year 5 mean = 12.66) obtained significantly higher scores compared with the younger children (Year 3 mean = 11.07). As before, there was also a significant main effect of modality ($F(1,67) = 5.74$, $p = 0.019$, $\eta^2_p = .079$), because scores for the reading modality (mean = 12.56) were significantly higher than listening modality (mean = 11.17). In contrast to the question answering data, the main effect of presentation format was significant ($F(1,67) = 4.95$, $p = 0.029$, $\eta^2_p = .069$): higher scores were obtained for story retells in the whole (mean = 12.31) compared to segmented format (mean = 11.41). In contrast to the question answering data, the main effect of inference type was not significant ($F(1,67) = 2.46$, $p = 0.122$, $\eta^2_p = .035$): mean scores for the local and global coherence questions were 11.56 and 12.17 respectively. Critically, the same two-way interaction reported for the question answering data was also evident: inference type and modality ($F(1,67) = 13.85$, $p < 0.001$, $\eta^2_p = .171$). All other interactions were not significant (all Fs < 1.47, ps > .05). The significant interaction between inference type and modality is depicted in Figure 2 and the same pattern as for the question-answering data is evident**;** post hoc t-tests indicated that the interaction arose

because performance on the global coherence inferences did not differ between modalities (t(69) = -0.01, p = 0.992), but did differ for local coherence inferences (t(69) = -3.65, p = 0.001; threshold for multiple comparisons = 0.0125), for which scores were significantly greater for the reading (mean = 26.11) compared to the listening modality (mean = 20.36).

Table 3 around here

Figure 2 around here

Discussion

The present study investigated typically developing children's ability to generate inferences to establish local and global coherence assessed by comprehension questions and story retells. To understand better the conditions under which children make coherence inferences, we contrasted whole and segmented story presentation formats for all participants, and reading and listening comprehension between groups. Both the comprehension questions and story retells were sensitive to developmental differences, with better performance for older than younger children, and also to modality differences with better performance for reading than for listening comprehension. There was an interaction between inference type and modality for both comprehension measures. This arose because scores for the local coherence inferences were relatively low in the listening compared to the reading modality. In contrast, there was no difference between modalities on the global coherence inferences. Of note, scores for local coherence in the listening modality were lower than scores for global coherence in the listening modality and both inference types in the reading modality. Main effects of presentation format and inference type were not consistent across comprehension measures. We discuss these findings in turn and their implications for our understanding of language comprehension and assessment as well as the clinical implications for children with comprehension difficulties.

Our findings demonstrate that both age groups were capable of making necessary coherence inferences to understand short narrative texts. This is consistent with previous research that has demonstrated children are able to resolve local and global coherence inferences in order to construct a mental model of the story (e.g. Ackerman, 1988; Currie & Cain, 2015; Lynch et al., 2008). As expected, performance was significantly better for older children and this medium size effect (comprehension questions η^2_p = .295, story retells η^2_p = .101; Cohen, 1969) was apparent across both comprehension measures. Therefore our assessment of reading and listening comprehension by open-ended questions and retell is sensitive to developmental progression. Of note, the developmental difference did not interact with modality or presentation format suggesting that the main effect of age group reflects general development in language comprehension and narrative skill.

There was a significant main effect of inference type for the comprehension questions, representing a small size effect (η^2_p = .098, Cohen, 1969). Scores were higher for the global coherence questions compared to the local coherence questions. This is in line with the findings of previous studies that have used the same (Currie & Cain, 2015; Freed & Cain, 2017) and different (Carlson et al., 2014) materials.

There was no significant main effect of inference type for the story retells, although the pattern of data is still the same as for the comprehension questions, with higher scores for the global coherence aspects of the retell compared to the local coherence elements. We believe these findings may be due to the interdependent nature of the two types of inference in our task: both were required for the same stories and during construction of the mental model, the generation of inferences early in the narrative will impact upon the ability to interpret later information. Recent research into the dimensionality of inference making suggests that it may not be possible to separate fully the ability to generate local

and global coherence inferences from the same text if both are necessary and, therefore, required to construct a coherent mental model (Language and Reading Research Consortium & Muijselaar, 2018). It may be important to assess both local and global coherence inferences to obtain a comprehensive understanding of a child's ability and any difficulties as each inference type draws on different information (within the text vs general knowledge, which is outside of the text), as well as serving different functions (as discussed in more detail below).

For both comprehension measures (comprehension questions and retells) there was a significant effect of modality, which was small to medium in terms of effect size (comprehension questions η^2_p = .170, story retells η^2_p = .079; Cohen, 1969): children made more inferences when they read the stories themselves, than when listening to them. The effect of modality was qualified by a medium-sized interaction (comprehension questions η^2_p = .188, story retells η^2_p = .171; Cohen, 1969) with inference type, evident in the analysis of both the comprehension questions and story retells. Scores were significantly greater for local coherence inferences in the reading compared to the listening modality, but performance on the global coherence inferences did not differ significantly between modalities. To understand this interaction, we need to consider the different processing requirements of reading and listening modalities, as well as the different functions of global and local coherence inferences in our materials. Considering first the materials, Freed and Cain (2017) argued that for the same texts the global coherence inferences were more linked to central details in the story, needed to understand character's actions, settings, and key events, whereas the local coherence inferences supported comprehension of more peripheral details (Currie & Cain, 2015; Miller & Keenan, 2009). This difference may underlie the difference in performance between the two inference types.

Turning to modality, previous empirical studies with different materials and participants and using a within-group design have reported an advantage for reading over listening comprehension (Cain & Bignell, 2014; Language and Reading Research Consortium et al., 2018). In some respects this advantage for reading is surprising because processing of written text requires decoding, which is a significant limiter of reading comprehension in beginner readers who are learning to read (Hoover & Gough, 1990; Language and Reading Research Consortium, 2015b; Perfetti, 1985). In general, listening comprehension becomes the more significant predictor of reading comprehension as children develop word reading fluency and transition to reading to learn, which occurs between 8 to 10 years of age (Garcia & Cain, 2014; Language and Reading Research Consortium, 2015b), the age of the children in this study. In other research that has also found an advantage for reading over listening comprehension, researchers have postulated that reading may enhance comprehension because it helps focus attention on the text, enables the participant to control the pace of information delivery, and also enables re-reading; for listening tasks, material cannot be processed if missed due to a lapse in attention (e.g., Cain & Bignell, 2014; Language and Reading Research Consortium et al., 2018). Further, poor comprehenders are better able to make inferences in response to open ended questions when the text is available for review (Cain & Oakhill, 1999). Together, these features may explain why children performed better on the less critical text details (local coherence inferences) in the reading than in the listening condition in our study: reading the text supported attention and enabled re-reading or checking of the text to support inference making. Future research could systematically vary local and global coherence inferences in relation to central and peripheral events in a text as discussed above (Miller & Keenan 2009) to understand better

the conditions that support inference making in the service of constructing a mental model of a text's meaning.

It is important to note that different children completed each task modality. However, participants were randomly assigned to the assessment modality and the groups did not differ significantly between modalities in age or vocabulary. This limitation of our study design was dictated by the need to minimise session length and time out of classroom lessons, as well as by our pool of previously tested items. However, we note that the advantage of reading over listening comprehension has been found for different groups of children using other materials in studies using a within-group design (Cain & Bignell, 2014; Language and Reading Research Consortium et al., 2018). Thus, we do not think that our findings were simply the result of subtle undetected differences between the groups who completed the tasks in each modality. To test the source of the reading comprehension advantage, future studies could compare performance on reading and listening comprehension tasks in the same groups of children and, in addition, examine the extent to which variation in memory and attention, and also background knowledge, influence performance.

We did not find a statistically significant benefit of one presentation format (whole vs segmented) over the other for the comprehension questions. The addition of the retell task may have had an impact here and we discuss this further in the limitations. In contrast, for story retells children's scores were statistically higher after the whole presentation of the story, compared to segmented presentation. However, we note that the size of the effect was small (η^2_p = .069; Cohen, 1969). van den Broek, Tzeng, et al. (2001) also reported poorer story retells for children who were asked questions during text reading. Together, these findings suggest that whilst segmentation may help children to make inferences whilst

processing a story (see also Freed and Cain, 2017), this information may not have been encoded into the mental model of the text that they were constructing and, as a result, was not included in their retells. This finding has potential educational and clinical relevance as 'online' questioning during reading is frequently used to develop reading comprehension skills (Department for Education, 2014) and is used as a facilitative technique by speech and language therapists (Camilleri & Law, 2007; Hasson et al., 2012; Hasson & Joffe, 2007). If this approach does not enhance story retell ability then it is important that clinicians and educators are aware of this when working on narrative skills. This needs further exploration.

Our final aim was to test the reproducibility of the findings of Freed and Cain (2017) for the comprehension questions. In line with Freed and Cain (2017), Year 5 children did better overall than those in Year 3. We also found that scores were higher for the global coherence questions compared to the local coherence questions overall and that there was no overall effect of presentation format. Whilst we did not find a significant interaction between inference type and year group in this study, this interaction approached significance and the pattern of means for the listening task was similar for the children in both studies, with a more substantial increase in scores for the local coherence questions with increasing age, compared to a smaller increase for global coherence questions. In contrast, this pattern is not present in the reading data. This difference in modalities may explain the non-significant interaction between inference type and year group (and therefore the significant interaction between modality and inference type). Freed and Cain (2017) also reported a clear benefit of segmentation for the Year 3 children, particularly for the local coherence inferences. The data in the current study also shows that for the Year 3 children in the listening condition, segmentation does appear to advantage local coherence. Thus, although we have not reproduced that finding in terms of a significant interaction, the

pattern of means fits with previous work. Segmentation also appears to benefit the Year 3 children who completed the reading modality suggesting that segmentation may be helpful for answering comprehension questions across modalities for younger children.

Overall, despite the methodological differences between the two studies, it is a strength of these studies that the findings in terms of effects of age and inference type and the pattern of findings in relation to presentation format were reproduced with a different sample of children and to some degree in a different modality of task.

Limitations

In addition to the limitations already discussed, we note the following. First, story retells were not independent of the comprehension questions because the questions were always asked before the retell (either in segmented or whole presentation format). Questions focused the child's attention on the key points in the story, prompted them to make an inference, and through that may have directed attention to important information to be included in the retell (Silva & Cain, 2019). Of note, retells were poorer in the segmented format condition. Processing the questions in sequence immediately after the story presentation may have enhanced the mental model and subsequent retell in the whole presentation condition. Other research suggests that questions asked before presentation of the story can usefully direct attention to core story features (Silva & Cain, 2019). This needs further exploration with both typically developing and clinical populations, to determine how the timing of questions can support narrative comprehension. In relation to this point, we did not fully reproduce the findings of Freed and Cain (2017) who reported an interaction between inference type and year group; here we found the same pattern of findings, but the interaction was not statistically significant because the age group differences were not as pronounced. We speculate that for some children story retell for

the practice story, may have alerted them to a failure to comprehend if they included few, if any details, in their retell. As a result, those children may have adjusted their strategy for processing the text in the target experimental stories that followed, in order to meet the task demands (van den Broek, Lorch, et al., 2001). If so, we may have unwittingly influenced their standard of coherence for the subsequent test stories and reduced differences between presentation formats in this study.

Second, local and global coherence inferences each involve access to relevant vocabulary and background knowledge. Although groups were matched for receptive vocabulary, we did not assess child's background knowledge when a child failed to make the target inference. Such checks have been used successfully in previous research to rule out lack of knowledge as a source of younger children's and poor comprehenders' failure to make necessary inferences (Barnes, Dennis, & Haefele-Kalvaitis, 1996; Cain, Oakhill, Barnes, & Bryant, 2001). However, a variety of child characteristics may influence the generation of coherence inferences. Therefore, we do not believe that a lack of vocabulary or background knowledge provides the sole explanation for failure to generate inferences and for developmental and ability differences: when knowledge is controlled, younger children and poorer comprehenders still make fewer coherence inferences than older and stronger comprehenders (Barnes et al., 1996; Cain et al., 2001). Finally, whilst we observed developmental changes, we tested only two age groups. It would be useful in the future to include a broader age range. In particular it would be important to explore the effect of segmented presentation for younger children who are beginning to read and therefore have limited decoding skills. This would allow exploration of whether the segmented format is facilitative for comprehension at this stage of reading development or is burdensome given younger children's poorer decoding skills.

Implications for practice

Our results are important for clinicians, teachers and other professionals assessing and supporting comprehension skills for the following reasons. It is vital to identify children with poor comprehension in order to develop appropriate support strategies and limit the adverse impact of poor comprehension on educational attainment. The results from this study suggest it may be useful to consider the modality and presentation format for comprehension tasks so that areas of strength can be utilised to facilitate effective comprehension. More information may be needed at an individual level about other related skills, for example, working memory and attention, vocabulary and background knowledge, and also word decoding skills, to identify the best way to support each child. For example, for children who have good decoding ability and those who have attention or memory difficulties, it might be more beneficial to use the reading rather than listening modality as this reduces memory and attention demands, and they can refer back to the text if needed.

The optimum form of input and structure of the stimuli may also depend on the aims of a specific task or the skill that is being assessed as we found some differences in the pattern of results for the two comprehension measures. The results suggest that the segmented approach may not be beneficial for story retells. Therefore if a child is focussing on developing their narrative skills, questions may be useful to help the child to structure their retell, but whole rather than segmented presentation format may be more appropriate to facilitate retelling. Most of the comprehension questions for the stories in this study were Wh questions. These have been successfully used in interventions to improve the quality of narratives in children with delayed language development (e.g. Davies et al., 2004). This requires further exploration.

Comprehension questions may support the child by focusing attention to key details, allowing them to construct a more accurate and integrated mental representation of the text. Critically, targeted questions may identify a child's potential to generate an inference when asked in segmented format. This may assist teachers to identify children who may require more targeted or specialist intervention from a speech and language therapist. For comprehension questions, our results, and those of Freed and Cain (2017), suggest that using questions in the segmented format may be useful for younger children or those who need support to develop their inference skills for comprehension questions. This also fits with advice to use 'online' questioning for development of comprehension skills in UK National Curriculum (Department for Education, 2014) guidelines, and dynamic assessment approaches used by speech and language therapists with children who have language disorders (Camilleri & Law, 2007; Hasson et al., 2012; Hasson & Joffe, 2007). For these children, supporting them with these fundamental inference skills, to make links between different parts of the story or to incorporate background knowledge, may be a higher priority than being able to produce a coherent story retell. Therefore the strategies and assessment methods used may depend on the learning outcomes for the specific situation as well as for the individual child. We note that most standardised assessments of reading and listening comprehension used to measure reading ability and diagnose difficulties (e.g. NARA; Neale, 1997; YARC; Snowling et al., 2009; WIAT; Wechsler, 2005) assess comprehension with questions at the end of the text, as in our whole presentation format, and are further limited by a lack of systematic contrast of inference and literal detail. Future research should explore the use of segmented questions to establish potential for inference making, and also to encourage deeper encoding of material and promote comprehension.

The results from this study support other recent research indicating an advantage for reading over listening comprehension, a finding that has important implications for classroom practice and the development of appropriate assessments. The current study tested typically developing children. It is important for this work to be expanded to clinical populations to see if the same pattern of results is observed in children with language difficulties. It may be important for speech and language therapists working with these groups to consider that comprehension in the reading modality may be better than in the listening modality. If this is the case, the reading modality may provide a means of support or a useful intervention strategy for these children. If reading can be used to support development of verbal comprehension this could be a useful therapeutic device and may be utilised in the classroom. The findings therefore have implications for the design of comprehension assessments as well as for supporting children with comprehension difficulties in the classroom.

References

Ackerman, B. P. (1986). Referential and causal coherence in the story comprehension of children and adults. *Journal of Experimental Child Psychology, 41*(2), 336-366. doi:10.1016/0022-0965(86)90044-5

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

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*(3), 216-241. doi:10.1006/jecp.1996.0015

Bishop, D. V., & Adams, C. (1992). Comprehension problems in children with specific language impairment: literal and inferential meaning. *Journal of Speech & Hearing Research, 35*(1), 119-129.

Cain, K., & Bignell, S. (2014). Reading and listening comprehension and their relation to inattention and hyperactivity. *British Journal of Educational Psychology, 84*(1), 108-124. Retrieved from <Go to ISI>://WOS:000331386800006

Cain, K., & Oakhill, J. (1999). Inference making ability and its relation to comprehension failure in young children. *Reading and Writing, 11*(5-6), 489-503. Retrieved from <Go to ISI>://WOS:000084003900004

Cain, K., Oakhill, J., & Bryant, P. (2000). Investigating the causes of reading comprehension failure: The comprehension-age match design. *Reading and Writing, 12*(1-2), 31-40. doi:10.1023/a:1008058319399

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.

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*(4), 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*(3), 258-274. doi:10.1080/1034912x.2014.934004

Casteel, M. A. (1993). Effects of inference necessity and reading goal on childrens inferential generation. *Developmental Psychology, 29*(2), 346-357. doi:10.1037//0012-1649.29.2.346

Cataldo, M. G., & Oakhill, J. (2000). Why are poor comprehenders inefficient searchers? An investigation into the effects of text representation and spatial memory on the ability to locate information in text. *Journal of Educational Psychology, 92*(4), 791-799. doi:10.1037//0022-0663.92.4.791

Catts, H. W., Compton, D., Tomblin, J. B., & Bridges, M. S. (2012). Prevalence and nature of late-emerging poor readers. *Journal of Educational Psychology, 104*(1), 166-181. https://doi.org/10.1037/a0025323

Cohen, J. (1969).Statistical power analysis for the behavioral sciences. New York: Academic Press.

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.

Davies, P., Shanks, B., & Davies, K. (2004) Improving narrative skills in young children with delayed language development, Educational Review, 56:3, 271-286

DCSF. (2008). *The Bercow Report: A review of services for children and young people (0-19) with speech, language and communication needs*. London: Department for Children Schools and Families.

Department for Education. (2014). Statutory guidance. National curriculum in England: English programmes of study. Retrieved from www.gov.uk/government/publications/national-

curriculum-in-england-english-programmes-of-study/national-curriculum-in-england-english-programmes-of-study

Dunn, L. M., Dunn, L. M., Whetton, C., & Burley, J. (1997). *British Picture Vocabulary Scale II*. Windsor: NFER-Nelson.

Elwér, Å., Gustafson, S., Byrne, B., Olson, R. K., Keenan, J. M., & Samuelsson, S. (2015). A retrospective longitudinal study of cognitive and language skills in poor reading comprehension. *Scandinavian Journal of Psychology*, 56, 157-166. https://doi.org/10.3102/0034654313499616

Freed, J., & Cain, K. (2017). Assessing school-aged children's inference-making: the effect of story test format in listening comprehension. *International Journal of Language & Communication Disorders, 52*(1), 95-105. doi:10.1111/1460-6984.12260

Garcia, J. R., & Cain, K. (2014). Decoding and reading comprehension: a meta-analysis to identify which reader and assessment characteristics influence the strength of the relationship in English. *Review of Educational Research*, 84(1), 74-111 Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101*(3), 371-395. doi:10.1037//0033-295x.101.3.371

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*(3), 285-299. doi:10.1111/j.1460-6984.2011.00108.x

Hasson, N., & Joffe, V. (2007). The case for dynamic assessment in speech and language therapy. *Child Language Teaching and Therapy, 23*(1), 9-25.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing, 2*(2), 127-160. doi:10.1007/BF00401799

Hulme, C., & Snowling, M. J. (2011). Children's Reading Comprehension Difficulties: Nature, Causes, and Treatments. *Current Directions in Psychological Science*, 20(3), 139-142. https://doi.org/10.1177/0963721411408673

Keenan, J. M., Hua, A. N., Meenan, C. E., Pennington, B. F., Willcutt, E., & Olson, R. K. (2014). Issues in identifying poor comprehenders. *LAnnee psychologique*, 114, 753-777.

Kendeou, P., van den Broek, P., White, M., & Lynch, J. S. (2009). Predicting reading comprehension in early elementary school: the independent contributions of oral language and decoding skills. *Journal of Educational Psychology*, 101, 765-778. doi: 10.1037/a0015956

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York, NY: Cambridge University Press.

Landis, J. R., & Koch, G. G. (1977). Measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174. doi:10.2307/2529310

Language and Reading Research Consortium. (2015a). The dimensionality of language ability in young children. *Child Development, 86*, 1948-1965.

Language and Reading Research Consortium. (2015b). Learning to read: should we keep things simple? *Reading Research Quarterly, 50*, 151-169. doi:10.1002/rrq.99

Language and Reading Research Consortium, Jiang, H., & Farquharson, K. (2018). Are working memory and behavioral attention equally important for both reading and listening comprehension? *Reading and Writing, 31*, 1449-1477.

Language and Reading Research Consortium, & Muijselaar, M. M. L. (2018). The dimensionality of inference making: Are local and global inferences distinguishable? *Scientific Studies of Reading, 22*(2), 117-136.

Lepola, J., Lynch, J. S., Laakkonen, E., Silven, M., & Niemi, P. (2012). The role of inference making and other language skills in the development of narrative listening. *Reading Research Quarterly, 47*, 259-282. doi: 10.1002/RRQ.020

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. doi:10.1080/02702710802165416

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*(2), 99-113. doi:10.1007/s11881-009-0025-x

Neale, M. D. (1997). *Neale Analysis of Reading Ability II: Second Revised British Edition*: Windsor, England: NFER-Nelson.

Olson, G. M., Duffy, S. A., & Mack, R. L. (1985). Question-asking as a component of text comprehension. In A. C. Graesser & J. B. Black (Eds.), *The psychology of questions* (pp. 219-226). Mahwah, NJ: Erlbaum.

Perfetti, C. A. (1985). *Reading Ability*. New York: Oxford University Press.

Rust, J., Golombok, S., & Trickey, G. (1993). *Wechsler Objective Reading Dimensions*: Sidcup, U.K.: Psychological Corporation.

Schmidt, C. R., & Paris, S. G. (1983). Childrens use of successive clues to generate and monitor inferences. *Child Development, 54*(3), 742-759. Retrieved from <Go to ISI>://WOS:A1983QR24000023

Silva, M., & Cain, K. (2019). The use of questions to scaffold narrative coherence and cohesion. Journal of Research in Reading, 42, 1-17.

Snowling, M. J., Stothard, S., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E., . . . Hulme, C. (2009). *York Assessment of Reading for Comprehension (YARC)*: GL assessment.

van den Broek, P., Lorch, R. F., Linderholm, T., & Gustafson, M. (2001). The effects of readers' goals on inference generation and memory for texts. *Memory & Cognition, 29*(8), 1081-1087. Retrieved from <Go to ISI>://WOS:000174016100002

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 R. F. Lorch & E. J. O'Brien (Eds.), *Sources of coherence in reading* (pp. 353-373). Hillsdale, NJ: Lawrence Erlbaum.

van den Broek, P., Tzeng, Y., Risden, K., Trabasso, T., & Basche, P. (2001). Inferential questioning: Effects on comprehension of narrative texts as a function of grade and timing. *Journal of Educational Psychology, 93*(3), 521-529. Retrieved from <Go to ISI>://WOS:000170829400007

Wechsler, D. (2005). *Wechsler Individual Achievement Test. Second UK Edition*. London: Harcourt Assessment.

Appendix A. Example story with comprehension questions (Freed & Cain, 2017). 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 (Freed & Cain, 2017).

	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, standard deviation and range for age, BPVS, WORD and number of word reading errors for children completing listening and reading modality of task in Year 3 and Year 5.

| | Reading | | | | | | Listening | | | | | |
| | Year 3 (n=16) | | | Year 5 (n=20) | | | Year 3 (n=17) | | | Year 5 (n=20) | | |
	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range
Age (years;months)	8;3	3.34 months	7;10-8;9	10;3	3.48 months	9;10-10;9	8;4	3.67 months	7;10-8;9	10;3	3.19 months	9;9-10;9
BPVS standard score[a]	105.56	13.91	83-139	106.55	10.87	88-126	102.94	10.61	85-119	104.90	12.33	84-131
WORD standard score[ab]	108.00	17.90	83-136	106.60	11.50	80-123						
Number of word reading errors on reading comprehension task[b]	18.06	34.25	0-124	1.45	3.41	0-15						

Notes: [a]Standard scores (mean = 100, SD = 15). [b]Completed by children doing reading modality only. [c]Total number of words for six stories = 940

Table 2. Mean scores (standard deviation) out of 24 on whole and segmented versions of the comprehension questions, split by modality, year group and scores for the local and global questions.

	Year 3 (n=33)		Year 5 (n=40)	
	Local	**Global**	**Local**	**Global**
Reading				
Whole	17.06 (3.99)	17.31 (3.70)	21.10 (2.40)	20.20 (2.40)
Segmented	18.81 (2.83)	18.31 (3.03)	20.25 (3.14)	19.90 (2.27)
Listening				
Whole	13.82 (5.15)	17.06 (3.70)	18.40 (2.96)	19.00 (2.55)
Segmented	14.88 (2.89)	17.41 (2.00)	17.85 (2.60)	19.50 (2.46)

Table 3. Mean scores (standard deviation) out of 24 for story retell based on the whole and segmented versions of the task, split by modality, year group and scores for the local and global questions.

	Year 3 (n=33)		Year 5 (n=38)	
	Local	**Global**	**Local**	**Global**
Reading				
Whole	12.50 (4.87)	11.19 (2.83)	14.16 (3.99)	13.58 (2.63)
Segmented	11.56 (4.29)	11.06 (5.04)	13.68 (3.09)	12.74 (2.81)
Listening				
Whole	10.35 (4.37)	12.47 (2.94)	11.68 (4.10)	12.58 (2.99)
Segmented	8.59 (3.32)	10.82 (3.50)	9.95 (4.02)	12.89 (2.85)

Figure 1: Graph depicting the two-way interaction between modality and inference type on the comprehension questions (scored out of 48).
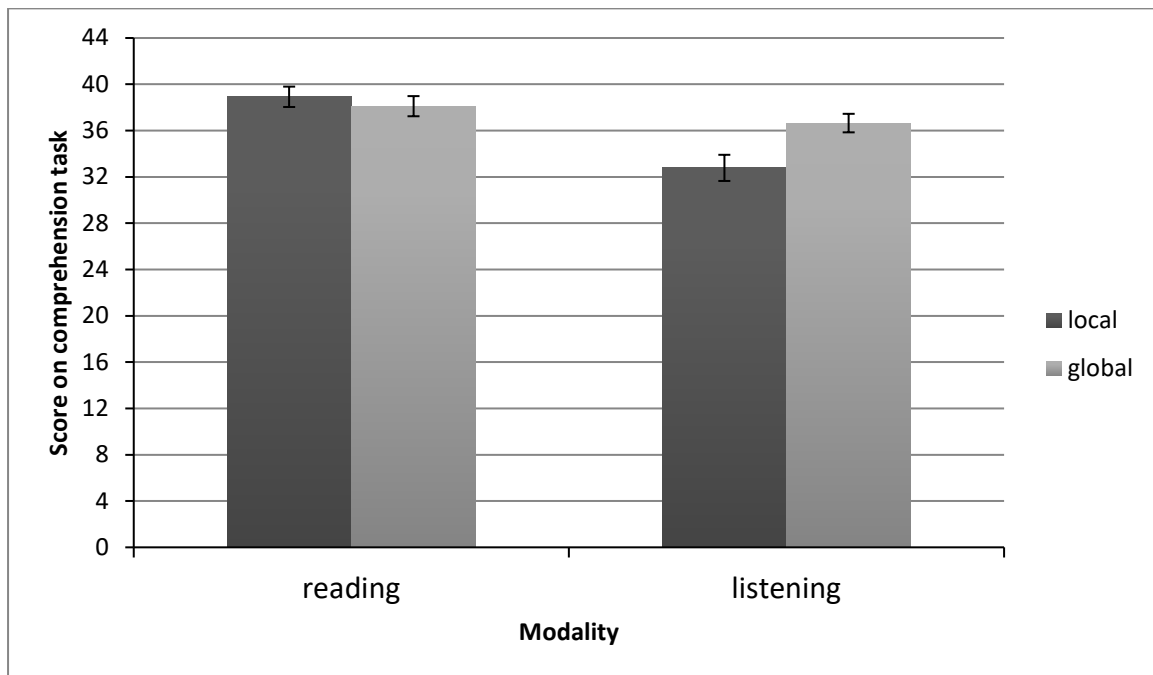
Figure 2: Graph depicting the two-way interaction between modality and inference type on the story retell (scored out of 48).