An Investigation into the Contribution of Inferencing in Children's Reading Comprehension

Kim R. Hewett

Bachelor of Psychology (Honours)

Murdoch University

School of Psychology and Exercise Science

Supervisor:

Dr. Bethanie Gouldthorp

This thesis is presented in partial fulfilment of the requirements for the degree of Bachelor of Psychology with Honours, Murdoch University.

October 2014

I declare that this thesis is my own account of my research and contains as its main content work which has not previously been submitted for a degree at any tertiary educational institution.

Kim R. Hewett

Copyright Acknowledgement

Signed:	____________________________
Author:	Kim R. Hewett
Full name of Degree:	Bachelor of Psychology with Honours
Thesis Title:	An Investigation into the Contribution of Inferencing in Children's Reading Comprehension
Year:	2014

Table of Contents

Acknowledgements

To Dr. Bethanie Gouldthorp

I would like to express my sincere appreciation for all of your time, effort and support with this research and thesis. Your guidance and supervision, as well as our open and honest communication, has been wonderful. It was such a pleasure working with you. Thank you!

To the staff, parents of students, and students at Perth public primary schools

Thank you all so much for your time, accommodation, facilities and support, allowing us to carry out data collection.

To Maroulia Katsipis

Thank you for your dedication to this project, your time and efforts are greatly appreciated.

To Cara Mueller

Your understanding and friendship have helped keep me sane during this hectic time. Thank you for being there.

To my friends and family

Your unwavering support and confidence in my ability mean so much to me. Thank you all for being so patient and understanding. Special thanks to Jimi, for always having confidence and believing in me (even when I sometimes found it hard to believe in myself) and for supporting me in every way throughout this entire degree. I will be forever appreciative.

Abstract

To date, little research has investigated the involvement of higher-level cognitive processes in reading comprehension ability in children. This study therefore aimed to help fill this gap in the literature by investigating the contribution of inferencing ability: a higher-level cognitive process which is important for successful reading comprehension. To measure inferencing, 38 children aged eight to 11 years participated in a computer task, which measured reaction time and accuracy. Participants were presented with narratives followed by word or non-word targets, on which they made a lexical decision. Word targets were either inference neutral or inference related, with inference generation measured as the difference between the means of the two. Participants were first categorised into low and high comprehension groups by an innovative higher-level comprehension measure, and again by a standardised, lower-level comprehension measure, and then compared. As the higher-level measure purports to directly measure skills underlying inference generation, a finer distinction in inferencing ability was expected between groups when measured this way, compared to the lower-level measure. Overall, however, the data did not reflect whether inferences were generated or not, nor meaningful categorisation between groups, irrespective of which measure they were categorised by. Interpreting the extent of inference generation between high and low comprehenders was therefore difficult. It is likely that the inferencing measure was unable to identify inference generation and that this, along with some other methodological limitations, accounts for the lack of differences between low and high comprehenders. Nevertheless, this study is an important contribution to the literature, as it provides a foundation on which future investigation can be based by illuminating methodological issues which first need to be addressed.

An Investigation into the Contribution of Inferencing

in Children's Reading Comprehension

In civilised societies, the ability to read and comprehend a text are fundamental skills, pertinent for success in academia, communication, many forms of employment, acquiring and applying knowledge, and also for the personal enjoyment of experiencing fictional worlds (Cain, 2009). Yet, reading can be a challenging task, especially for complex, technical, or unfamiliar material. Moreover, for some individuals, reading is challenging all of the time (McNamara, 2007). Research has demonstrated this complex nature of the task of reading, however successful reading can broadly be conceptualised as the acquisition and integration of two sets of skills: those enabling word decoding, and those supporting the comprehension of meaning (Cain, 2009; Cain & Oakhill, 2008). Comprehension can be defined as the ability to go beyond the words in a text, to understand the meaning of ideas presented in the text and to determine the relationships between them (McNamara, 2007).

A vast amount of research has attempted to determine the precise skills and cognitive processes that underlie reading comprehension, documenting the difficulties which can arise as a result of deficiencies in these areas. However, despite enormous efforts to promote successful reading comprehension, approximately 10% of school children have age-appropriate word reading skills, and yet have poor reading comprehension (Nation, Clarke, Marshall, & Durand, 2004). To help alleviate this issue, it is crucial that a better understanding of the underlying processes involved in children's reading comprehension ability is achieved. Moreover, before poor comprehenders can be helped, researchers need to determine which of the many factors of comprehension skill are causally implicated in reading comprehension development (Cain & Oakhill, 2007). The majority of research to date has focused on skills that have consistently been found to be highly predictive of comprehension ability, such as word decoding (e.g. Gough & Tunmer, 1986; Perfetti & Hart, 2001; Perfetti & Hogaboam, 1975; Shankweiler, 1989), spelling (e.g. Shankweiler et al., 1999) and phonological awareness (de Jong & van der Leij, 2002). These skills have consistently been associated with efficient word reading, and are termed “lower-level skills”, as they are required at the early stages of reading comprehension (Perfetti, 1985). Comprehension is aided by proficiency in lower-level skills; however, it is the integration of “higher-level skills” (e.g. inference-making, comprehension monitoring) that is required for a richer construction of meaning (Cain & Oakhill, 2008; Oakhill, Cain, & Bryant, 2003; Perfetti, 1985). When readers struggle with lower-level processing skills, deeper comprehension can easily be compromised, irrespective of higher-level processing ability. This is because labouring over lower-level skills can leave the reader with inadequate processing capacity to utilise higher-level processes, such as construction of a coherent, meaningful representation of the text (Perfetti, 1985).

Levels of Text Representation

The lowest level of text representation is the *surface form*, which refers to a representation of specific words and syntax that are used in a discourse (Van Dijk & Kintsch, 1983). Decoding occurs at this level, requiring the reader to recognise and retrieve the correct meanings of the individual words, which are then combined with syntactic form in order to construct meaningful sentences (Cain, 2009).

The middle level is the *propositional textbase*, an abstract representation of the text’s meaning, which is formed by linking and relating the ideas and concepts expressed in separate sentences together. By integrating the meanings of successive sentences, readers achieve local coherence (Cain, 2009; Van Dijk & Kintsch, 1983). Local coherence refers specifically to the coherence of individual segments of text adjacent to each other in the discourse. Local coherence is achieved when a small group of sentences makes sense alone, or with generally available knowledge. Therefore, if information from another section of that discourse is required in order for it to make sense, local coherence has not been achieved (McKoon & Ratcliff, 1992). However, integrating successive sentences is not always sufficient for understanding the meaning of concepts in a text. When processing information at a propositional level, the result is a representation solely of information explicitly presented in the text, limiting comprehension to literal meanings. If information is only processed at this level, difficulties will be encountered when comprehension also requires accessing implicit information; i.e., that which is not stated explicitly within the discourse, but is incorporated via the reader's world knowledge and/or prior experiences. Usually, in order to accurately comprehend a text, readers need to go beyond what is explicitly provided in that text, in order to establish how ideas fit together as a whole. This concept is referred to as global coherence (Cain, 2009; McKoon & Ratcliff, 1992).

Global coherence results in a representation of the situation described by the text, and here at the *discourse level*, the highest level of representation in language processing, the situation model formation is formed (Van Dijk & Kintsch, 1983). The lower-level surface form and propositional textbase mental representations merely form a description of the text itself. However, higher-level situation models are abstracted from the surface structure and textbase (Cain, 2009; Zwaan, 1999). Here, explicit information from the propositional representation is integrated with implicit information drawn from world knowledge, with the result being mental representations made up of extensive information of the spatiotemporal locations, individuals, settings, objects, causal relations between events, protagonist's goals, and actions (Cain & Oakhill, 2008; Van Dijk & Kintsch, 1983; Zwaan, 1999).

The differences between lower and higher-level processing were demonstrated by Zwaan (2004), using the following example: "The ranger saw the eagle in the sky." At the propositional textbase level of comprehension, only the explicitly expressed relations among entities, actions and locations are captured - a perceptual representation of the eagle is not accessed. By drawing on implicit world knowledge however, the reader enhances their situation model by inferring that since the eagle is in the sky, it is flying, and thus its wings will be outstretched. Zwaan, Stanfield, and Yaxley (2002) investigated responses to a picture of an eagle with its wings outstretched after reading either "The ranger saw the eagle in the sky" or "The ranger saw the eagle in the Tree". It was found that responses were faster to the picture after reading "The ranger saw the eagle in the sky", demonstrating the integration of perceptual information not contained within the text. This example shows that situation model formation produces a richer, higher-level form of comprehension. It is therefore important to understand how this higher-level representation is created.

Inferencing

Inferencing appears to be a crucial component of situation model formation, because this is the method used to draw implicit information into the representation. Inferencing is fundamental in ensuring a text is well understood, as writers do not always state every tiny detail, since this would result in a lengthy and potentially uninteresting text. Rather, the reader is required to fill in details not explicitly stated in the text, either by integrating explicit information from within the text or by incorporating world knowledge with textual information (Cain & Oakhill, 2007). Inferences are therefore used to derive a higher-level, overarching understanding of the state of affairs being described throughout a discourse (Cain & Oakhill, 2008; Gould, 2008). Two broad categories of inferencing exist: bridging and elaborative.

Bridging Inferences. Bridging (backward) inferences are those drawn on-line (during the reading process), when conceptual gaps occur in a discourse. An inference must be made on-line to fill that gap, in order to maintain coherence. This is done by connecting adjacent segments of text together, and helps to avoid misinterpretations or incomplete comprehension (Fincher-Kiefer, 1995; Gould, 2008; Graesser & Bertus, 1998). There is a general consensus concerning bridging inferences in the literature, which are suggested to be required for the construction of both local and global coherence (Fincher-Kiefer, 1995; Gould, 2008; Graesser & Bertus, 1998; Graesser, Singer, & Trabasso, 1994; McKoon & Ratcliff, 1992).

There are a few types of bridging inferences. Examples of anaphoric and causal antecedent inferences are detailed in the following Fincher-Kiefer (1995) narrative: "The salesman was sitting in the dining car of the train. The waitress brought a bowl of steaming soup to his table. Suddenly the train screeched to a halt. He jumped up and wiped off his pants." An anaphoric inference is required to understand that the pronoun "he" in the fourth sentence is the aforementioned salesman; and a causal antecedent inference is required to infer that the soup was spilled onto the salesman's lap (which is why the salesman was wiping off his pants), and that this was caused by the train stopping rapidly. Causal antecedent inferences refer back to the explicitly stated material in a discourse, but only in a way which attributes a cause for an action that has previously occurred. Lastly, "We got the party supplies out of the car. The ice-cream was melting." requires a connecting inference in order to link the two sentences, such that the reader infers that the ice-cream was one of the party supplies. These examples demonstrate that inferencing is required for successfully comprehending all but fairly simplistic discourse (Oakhill & Cain, 2000); highlighting the importance of understanding how inferencing ability contributes to reading comprehension, in order to advance both teaching and learning of reading comprehension.

Elaborative Inferences. Elaborative (forward) inferences embellish what is explicitly stated in a discourse, rather than explaining it (Gould, 2008; Graesser et al., 1994). Elaborative inferences are thought to be generated off-line, during later retrieval tasks (Graesser et al., 1994) and correspond to a deeper understanding of the text (Kintsch, 1988). Elaborative (forward) inferences are not essential to comprehension or coherence, they are considered optionally generated.

Several different types of elaborative inferences exist. For instance, inferring “spoon” from “Kim stirred her coffee,” requires an instrumental inference, integrating implicit knowledge from past experiences of usually stirring coffee with a spoon. Inferring “round” from “The man rolled an orange,” requires a semantic elaborative inference, integrating implicit knowledge that in order for something to roll, it must be round in shape, which oranges usually are. Lastly, predictive (causal consequence) inferences occur when the reader combines implicit knowledge with explicit text information, enabling them to generate expectations about what might happen next. For instance “Mitch the bulldog did not look very well cared for. The vet realised he was covered in fleas. Mitch lifted his back paw to up to his ear,” may elicit the predictive inference that Mitch was about to scratch, which may or may not be verified later in the text. Situation model formation is dynamic, meaning that as new information is presented, it is checked for coherence against previous implicit and explicit details, and the situation model is then updated (Schmalhofer, McDaniel, & Keefe, 2002). In the previous example, if Mitch did not go on to scratch his ear, the prediction that he was going to scratch would be discarded.

While both predictive and bridging inferences may be utilised to build a situation model, they have been found to do so at different levels of text representation (Fincher-Kiefer, 1993; Fincher-Kiefer & D'Agostino, 2004). The processes involved may differ substantially, even occurring in different areas of the brain (e.g. Beeman, Bowden, & Gernsbacher, 2000). Inferencing abilities develop in middle childhood, therefore, to determine age-related differences between bridging and predictive inferences, Barnes, Dennis, and Haefele-Kalvaitis, (1996) conducted a study on children aged six to 15 years. Children were trained on a novel knowledge base and then asked questions about a multi-episode story, requiring inferences which drew only on that knowledge base. It was found that the processing characteristics of bridging and elaborative inferences differed, such that: bridging inferences were made more frequently than elaborative inferences for all ages; and that memory for text was related to bridging but not elaborative inferencing. Results also showed that age-related differences in both bridging and elaborative inferencing did not weaken as a result of ensuring the knowledge base was equally available to all children, suggesting that inferencing ability develops with age, independently of the influence of knowledge.

Moreover, research has repeatedly shown that impairment in inferencing skill is related to children's reading comprehension difficulties (e.g. Cain & Oakhill, 1999; Cain, Oakhill, Barnes, & Bryant, 2001), such that those with poor reading comprehension skills are less likely to integrate meaning across sentences and combine text information with general knowledge to generate inferences, thereby impairing situation model construction (Oakhill & Cain, 2000). Further, inferences are only able to be generated if the reader has the requisite knowledge (Oakhill & Cain, 2000).

Assessing Comprehension

As previously detailed, success comprehending a discourse may be impacted by deficits at any level of text representation, in a number of areas. Deficits in any one of these areas can impact the development of situation models, inhibiting deeper comprehension (Francis et al., 2006). Reading comprehension assessment is therefore challenging, because multiple skill deficits may be present in the same individual, and thus can impact the results of a reading comprehension assessment. Further, separating the impact of lower-level and higher-level skills deficits is difficult, as lower-level deficits can constrain higher-level skills (Francis et al., 2006). The ability to uncover the precise areas of difficulty being experienced is paramount, as this in turn can lead to effective, targeted instruction for the specific difficulties that the individual is encountering (Francis et al., 2006; McNamara, 2007).

At present, in line with past research focus areas, tests of reading and comprehension ability currently utilised in schools focus on lower-level cognitive skills, and are not designed to provide diagnostic information (Francis et al., 2006; Spooner, Baddeley, & Gathercole, 2004). For instance, the Neale Analysis of Reading Ability (NARA-III; Neale, 1999) is a standardised measure currently utilised by schools in many areas of the world. This measure assesses “accuracy” (word decoding and fluency) and “comprehension”, which also predominantly taps into lower-level, propositional processes. Although word reading errors are corrected by the administrator during assessment, this is insufficient to measure comprehension completely independently of accuracy (Spooner et al., 2004). Labouring over pronounciation and then having to wait for a word to be supplied may also detract from the cohesion and global coherence of the text, potentially even leaving inadequate processing ability for higher-level processing to occur (Perfetti, 1985). Time delays may cause the reader to forget part or all of what they were reading previously from their working memory, therefore rendering them unable to integrate information from earlier in the text with the remainder of the text (Radvansky & Dijkstra, 2007). Moreover, assessment of higher-level skills is absent, despite these skills being critical for many facets of deeper comprehension, such as correct interpretation of non-literal language (e.g. metaphors), gaining an overarching understanding of an extended discourse, and generating global coherence.

It is therefore possible that some children may be struggling with reading comprehension as a consequence of higher-level processing difficulties, but because they can perform well on current lower-level reading comprehension assessments such as the NARA-III, these difficulties are not being detected early (Cain, 2009; Francis et al., 2006). Indeed, reading comprehension difficulties have been found in some children with normal decoding abilities (e.g. Oakhill & Cain, 2000; Oakhill, Cain, & Bryant, 2003), an issue which becomes apparent only when questions requiring more than recall of simple facts are asked (Cain, 2009). Further, recent evidence suggests that skills other than those that relate to lower-level reading processes may better predict children's comprehension level, such as comprehension monitoring, working memory and inference-making (e.g. Landi, 2010; Nation & Snowling, 1998; Oakhill, Cain, & Bryant, 2003; Perfetti, Landi, & Oakhill, 2005).

In support of these recent findings, a growing number of measures which tap into higher-level comprehension processes are being developed and researched. For instance, the Diagnostic Assessment of Reading Comprehension (DARC; August, Francis, & Calderón, 2002) is an innovative (though experimental) measure, which is claimed to assess higher-level processes pertinent for reading comprehension, without the requirement of concurrent lower-level processing skill. To score well in the NARA-III requires adequate reading accuracy and concept familiarity, which increases in complexity throughout the task. The DARC, however, controls for variation in vocabulary skill and background knowledge by incorporating only simplistic, well known words and concepts throughout, and training readers on a novel knowledge base. Comprehension questions related to different types of inferencing are then asked. Because inferencing depends on the availability and accessibility of the reader's knowledge base, by controlling for this amongst readers, the DARC seems to allow for more direct, independent measurement of the actual processes underlying inferencing (August, Francis, Hsu, & Snow, 2006).

Aims and Hypotheses

While the contribution of lower-level skills to reading comprehension have been researched at length, to date, much less research has been conducted on the contribution of higher-level skills. In order to target comprehension difficulties resulting from higher-level processing deficiencies, these processes and their contribution to reading comprehension ability need to be investigated and thoroughly understood. Accordingly, the current study seeks to help fill a gap in the literature by investigating the contribution of the higher-level skill of inferencing to reading comprehension level, a skill which has previously been suggested to be causally implicated in comprehension development (Cain & Oakhill, 1999, 2008).

In order to investigate this, the present study utilised a computer-based task to assess inferencing ability in children. Based on observations in past research that different inference types might be differentially related to comprehension level (e.g. (Barnes et al., 1996; Cain, Oakhill, Barnes, & Bryant, 2001; Gould, 2008), bridging and predictive inferences were included for comparison. Additionally, due to prior research suggesting that traditional measures of reading ability do not adequately capture higher-level comprehension ability (e.g. Francis et al., 2006; Spooner, Baddeley, & Gathercole, 2004), a comparison was made between the NARA-III (a traditional reading ability measure) and the DARC, which is claimed to assess higher-level processing. Participants will be categorised by each measure, enabling the identification of those children who have age-appropriate reading skills, yet struggle to comprehend what they're reading (low comprehenders) and comparing their results on the inferencing measure with results of children who have both age-appropriate reading and comprehension skills (high comprehenders). Both bridging and predictive inferencing ability will be analysed when comprehension is measured as a function of lower versus higher-level skills.

Because the DARC is purported to directly measure the processes underlying both bridging and predictive inferencing, it is expected to better distinguish low and high comprehenders by inferencing ability than the NARA-III. Therefore, it was firstly hypothesised that high comprehenders would have greater accuracy and faster reaction times for constructing both bridging and predictive inferences than the low comprehenders, when comprehension groups were generated using the DARC. As bridging inferences have found to be required for coherence, and predictive inferences are optional, it is expected that the difference between low and high comprehenders will be less pronounced for bridging than predictive inferences. The

NARA-III relies on sufficient reading skills and the availability of background knowledge. In line with related past research, it was therefore secondly hypothesised that when comprehension groups were generated using this measure, the high comprehension group would have greater accuracy and faster reaction times than the low comprehension group for construction of predictive, but not bridging inferences.

Method

Participants

Thirty-eight primary school students (21 female, 17 male) ranging in age from 8.25 to 10.83 years (*M* = 9.455 years, *SD* = 0.643) participated in this study, with two participants excluded due to non-compliance. This age range was selected as children should have sufficient reading abilities to participate in the tasks, while the restricted range should reduce confounds due to developmental level (Neale, 1999). Furthermore, inference skills have been shown to contribute to later comprehension skill at this age, more so than lower-level skills (Cain & Oakhill, 2008). All participants spoke English as their first language, had normal or corrected-normal vision, and were free from cognitive impairment. Students from participating classes were all eligible for the chance to win an iPod touch.

As socioeconomic status (SES) has been shown to be a predictor of comprehension abilities (Currie, 2008; Flouri & Buchanan, 2004; White, 1982), participants were recruited from schools ranging in SES to increase the probability of obtaining a wider range of comprehension abilities. The myschool.com.au website and the Australian Index of Community Socio-Educational Advantage (ICSEA) calculation were used to determine SES. The average value of all schools' ICSEA values is set at 1000. A total of 99 public schools were contacted (SES range 835-

1209), with five schools participating in this study, reflecting a range of low, moderate and high SES (909-1158).

Design

The present study employed a 2x3x2 mixed groups design, to manipulate the between groups independent variable (IV) of comprehension level (high or low) and the two within groups IVs of trial type (bridging, predictive or neutral) and target lexical status (word, non-word). Reaction time (RT) and accuracy were the dependent variables.

Materials

Neale Analysis of Reading Ability. The NARA-III (Neale, 1999) has been used extensively in both research and practice, and is age-appropriate for this sample. The version used was an adapted measure utilising a standardisation conducted with Australian students (Neale, 1999). It consists of two subtests: the first being a reading ability sub-test, which assesses word reading accuracy (contextual word recognition); and secondly, a lower-level reading comprehension sub-test. Accuracy (number of words read correctly, with higher scores indicating higher accuracy) and comprehension scores (number of correct answers, with higher scores indicating higher comprehension) were recorded. The accuracy sub-test score was used as a preliminary screening measure for any lower-level text processing issues, with those below 34 (year four) or 39 (year five) excluded from the study due to very low reading ability / possible reading disabilities (Neale, 1999).

The reliability and validity of this version were independently confirmed, with both found to be very good. High levels of internal consistency are reported by the developers, with reliability coefficients (KR-21) for year four of 0.94 for reading rate, 0.95 for reading accuracy, and 0.85 for reading comprehension; and for year five, 0.95 for reading rate, 0.96 for reading accuracy, and 0.96 for reading comprehension. Excellent retest reliability was also demonstrated, with high correlations between teacher and assessor administration (rate .95, accuracy .95 and comprehension .93). This indicates that using the manual, test administration remains consistent amongst individuals with varying levels of familiarity with the measure, therefore making the NARA-III appropriate for research purposes (Neale, 1999).

Concurrent validity was assessed for the NARA-III, with the developers reporting high correlations (.7 to .77) with the Dartmouth Advanced Reading Test on a sample of 200 students. Furthermore, a standardisation sample of over 1300 students showed progressive increases in score with age and years of schooling, providing evidence of construct validity. This measure was included to determine whether differences exist in inferencing ability for those classified as high and low comprehenders when comprehension is assessed as a function of low-level processing abilities.

Diagnostic Assessment of Reading Comprehension. The DARC (August et al., 2002) was adapted from Hannon and Daneman's (2001) university level measure, for use with primary school students. It was designed to measure individual differences in central comprehension processes, while minimising the need for high decoding ability or English oral proficiency levels.

Published reliability and validity information is currently minimal for this measure, however pilot studies assessing the DARC’s feasibility, developmental sensitivity, reliability, and relation to standardised measures have been conducted by August et al. (2006), with promising results. The first study involved 16 year two to year six students, and showed that the reading level of DARC items was at the appropriate level. A second study was conducted on 28 fourth graders who had scored poorly on the Woodcock-Johnson Language Proficiency Passages subtest, yet had a range of scores on the DARC. This demonstrated support for the discriminant validity of the measure, and indicated that yes-no answers were valid indicators of participant's thinking. The final pilot study, conducted with 521 students in kindergarten through to grade three, confirmed that the DARC was found to not be as strongly related to word reading as traditional measures, which may obscure children's comprehension capacities by requiring greater levels of syntax, decoding and vocabulary load. These findings provide preliminary support for the construct validity of the DARC.

A central proposition of this research is that the type of comprehension problems that result from higher-level processes may not be identified by traditional comprehension measures like the NARA-III. As the DARC has shown promise for tapping into higher-level processing skills independently of lower-level processing skills (August et al., 2006), this measure was included so that even if the NARA-III results are not related to inferencing ability, a relation may be demonstrated using the DARC, which would in turn provide evidence that traditional measures are not adequate for tapping into inferencing ability.

The DARC comprises four comprehension components: text inferencing (making inferences based on information provided in the text), knowledge access (recalling new text information from memory), text memory (accessing prior knowledge from long-term memory), and knowledge integration (integrating prior knowledge with new text information). The test was scored dichotomously, "1" for correct answers and "0" for incorrect, with higher scores indicating greater comprehension. The total range of possible scores is zero to 30.

Inferencing Measure. The inferencing measure was a computer-based lexical decision task, adapted from a task used previously in experimental settings to investigate inference generation (Fincher-Kiefer, 1993, 1995). Stimuli were modified to be age-appropriate and relevant for an Australian sample, and an additional 21 narratives were constructed to ensure a sufficient number of trials, following the same style and structure as the original stimuli. Narratives consisted of three sentences for the predictive and neutral conditions, three or four sentences for the filler condition, and four sentences for the bridging condition. These sentences together describe a scenario likely to generate either a bridging inference, a predictive inference, or no inference (neutral/filler). The fourth sentence in bridging inferences was included to complete the narrative, such that the reader could form a bridging inference by referring back to earlier in that narrative; whereas a fourth sentence in predictive inferences was omitted, in order to encourage the reader to complete the narrative by predicting the outcome.

Stimuli consisted of a total of 60 narratives that were followed by a word or non-word target. The word conditions (neutral, bridging or predictive) consisted of ten narratives each, and the remaining 30 narratives were the non-word condition (filler). Narratives were selected over other types of discourse (e.g., expository texts), as narratives have been shown to be more likely to elicit situation model construction. This occurs because narrative texts closely correspond to everyday experiences involving performing actions in pursuit of goals, goal obstacles, and emotional responses to these events (Graesser, Singer, & Trabasso, 1994). These similarities mean that narratives are more likely to elicit background information than expository texts, which are decontextualized and usually used to inform learning. Typically the reader will not have extensive background knowledge about topics in expository texts, and are therefore unable to generate as many inferences (Graesser et al., 1994). Sample narratives from each condition are shown in Table 1.

Table 1

Inferencing Stimuli Narrative Samples

Condition	Narrative	Target
Bridging	Jack stood on the platform high above the pool. As he looked down at the water below, he was nervous. He thought about all his practice and took a deep breath to calm himself. He felt great as he swam to the surface to see the score from the judges.	Dive
Predictive	The teenager went to the beach for Labour Day It was time to tan her pale skin. The sound of the water and the heat made her fall asleep in the sun.	Burn
Neutral	Georgia had not been allowed to go outside yesterday because it had been very windy. She hoped to be able to play with her friends outside this morning. As the sun came up, Georgia had the feeling she wasn't going to get what she wanted.	Snow
Filler (i.e., non-word trials)	Sarah put the cake on a plate and got the lighter ready. She lit the candles and brought the cake into the lounge. Everyone started singing "Happy Birthday". Some of the wax from the candles dripped onto the cake.	Bolze

Note. Narratives adapted from Fincher-Kiefer (1993, 1995). Refer to Appendix D for more samples.

The first three sentences of the bridging inference stimuli introduce the situation. The last sentence presents the result of the previous sentences. That Jack dived isn't explicitly stated, however it is possible to form that inference in order to understand the final sentence and accept this as a logical continuation of the situation (that is, the coherence of the text). The predictive inference that the teenager's skin burned when she fell asleep in the sun is likely to be generated, as this is easily accessible general knowledge. It is generally understood that staying in the sun for an unplanned and extended period of time when your skin is pale is likely to result in a burn. There are other possible outcomes (e.g. the teenager waking up shortly after falling asleep) but this is intuitively less likely, and is a less interesting outcome – which would elicit the question of what the point is of telling the story (Gould, 2008).

Target words in the neutral stimuli were chosen to be related to the narrative in order to control for relatedness priming effects, but not an expected outcome or inference for the explicit text. In this example, Georgia hadn't been able to go outside yesterday due to bad weather conditions (it was very windy). She now wanted to go outside to play, but had the feeling that she wasn't going to get what she wanted. From this narrative, a reader is unlikely to infer that the reason Georgia felt that she wasn't going to be allowed outside to play that day was because it was snowing. However, snow could be considered related to the story, because like excessive wind, snow could be considered poor weather conditions for playing outside. Targets in the filler stimuli were non-words in order to enable a lexical decision to be made.

Scoring of this measure involved the calculation of mean reaction times (RTs) for each participant's correct responses to target words for each condition. The means from each participant's inferencing trials were then subtracted from the mean of their neutral trial. This was done to assess the difference between the two inferencing conditions. Faster RTs in the inferencing conditions indicate that the participant was generating inferences, with greater difference scores indicating higher facilitation. Accuracy scores (number of targets correctly identified as words or non-words, with higher scores indicating higher accuracy) were also recorded. As data for the present study was collected in conjunction with another project; imagery, predictive inferencing (performed under a visuospatial and verbal load), intelligence, working memory, and temporal sequencing ability data were also collected, though are not relevant to the present study.

Apparatus

The inferencing measure was presented on a Toshiba Satellite C660 Intel 2.10 GHz laptop running Windows 7 operating system, with 2GB RAM. DirectRT (version 2010.2.103.1115) software was utilised to run the inferencing task at 1280x720 screen resolution, 32-bit colour at 85 hertz refresher rate. A DirectIn Button Box (Eternity+ Empirisoft Corporation, 2014) was connected to the laptop via USB. The button box had nine buttons on it: button one on the left side, to button nine on the right. Button one corresponded to a “yes” response and button nine corresponded to a “no” response, with each labelled as such to prevent confusion. To prevent participants accidentally pressing yes or no when the target word appeared, only buttons two through eight were used to bring up the next sentence in the narrative, not buttons one or nine. To prevent participants accidentally skipping comprehension questions, only letters on the laptop keyboard were able to be used to commence the next trial after comprehension questions were answered.

Procedure

Each participant was assessed at their school, in a quiet room separate from their classroom. Participants were tested individually, with an individual native English speaking researcher who was present throughout the entire session. Prior to commencing each measure, participants gave informed consent (see Appendix A) and were given the opportunity to ask any questions. Rapport was established with participants at the beginning of (and built upon throughout) each session. Participants completed the NARA-III (Neale, 1999) in session one, and then those who met inclusion criteria (NARA-III scores above 34 for year four, or 39 for year five) went on to complete the remaining two sessions. The DARC, inferencing measure, and additional measures not related to the current study (as previously noted) were completed in sessions two and three in randomised order, to reduce the likelihood of order effects. Task order in each session remained constant, however. The measures utilised in the present study were each completed in separate sessions, on different days, with each of the three full sessions taking approximately one hour. The NARA-III and inference measures were both completed at the start of the session, however, the DARC was not. Rest breaks were provided before each new measure was commenced, therefore fatigue effects due to task order were considered unlikely.

Neale Analysis of Reading Ability. The Australian standardised NARA-III manual (Neale, 1999) was followed step-by-step to administer the test. Participants read the test instructions, and were given the opportunity to ask the researcher any questions. Participants were instructed to first view the picture, and then read the passages aloud, with any word reading errors (e.g. mispronunciations, omissions) corrected throughout, and recorded on the participant’s individual record. Comprehension questions were asked immediately at the end of each passage. Correct answers were not given if the participant answered incorrectly.

Participants first completed practice passage Y, and once they understood what was expected, testing commenced. There are six form one passages, each increasing in difficulty. Scores were recorded by the researcher as the participant read. Participants continued through all six passages unless 16 reading accuracy errors were made for passages one to five, or 20 for passage six, at which point testing ceased and only scores recorded up to the previous passage were used.

Precautionary steps were taken to ensure no undue stress to participants if testing needed to be ceased, such as changing to a simpler passage to end on a note of success. Due to the potential for ambiguity of participant responses, when answers were given which were close to, but not exactly, the correct answer, consultation with a second scorer was carried out to ensure consistency of marking. This measure took approximately 20 minutes to complete.

Diagnostic Assessment of Reading Comprehension. Complete verbal instructions for this measure are provided in Appendix B. Participants read the practice passage aloud, so that the researcher could further explain the requirements of the task, and new (made-up) words as they were read. When the participant finished reading the passage it was removed from their view, and they were instructed to read the practice yes/no questions aloud one at a time and answer them. Participants were advised if their answer was correct or not, with the reasoning for the correct answer detailed. Explanations for incorrect answers were encouraged, with the researcher providing additional assistance to ensure the participant was able to arrive at the correct answer.

Once ready to complete the task, participants were advised to read passages silently thereafter. No additional aid was provided in response to questions, with first responses recorded. The first section of the task was read, covered, and five questions asked. Participants then re-read the first section, then the second section, with both sections covered and seven more questions asked upon completion. Finally, the first and second sections were re-read, and then the third section. All sections were then covered and a final 18 questions asked. This measure took approximately 15 minutes to complete.

Inferencing measure. Participants were instructed verbally throughout this measure (see Appendix C). Participants were seated at the laptop, and the researcher first entered the participant's demographic data including initials (identifier), gender, birthdate and school. Participants were then guided through reading four practice stories (two real words, two non-words), and given the opportunity for questions throughout and upon completion of the practice narratives.

Once the participant was ready, the task was commenced. All task stimuli appeared at the screen centre, in black font against a white background. Before each narrative, the words "get ready" appeared in size 20 Arial Black capitalised font for 2000 ms, to prepare participants to begin reading the next narrative. Each sentence of the narrative was presented individually, in size 24 Times New Roman font, and participants read at their own pace (as detailed earlier, the next sentence was presented when any of buttons two through eight were pressed by the participant). At the completion of each narrative, a fixation cross was shown for 1000 ms, to indicate that the target was about to appear (signalling that participants needed to get ready to make a decision by placing their hands over the response buttons).

The target was displayed in size 25 Arial font. Upon presentation of the target, participants made a lexical decision of whether it was a real word by pressing "yes" on the response box with their left hand, or not a real word by pressing "no" with their right hand. Although this is a go/go task, the non-word condition was included purely to enable the lexical decision task and thus responses to non-words are not of theoretical relevance to the study. Since no comparisons are to be made between RTs to word and non-word targets, counter-balancing of response hand was not necessary (i.e., the same response hand was used for all conditions on which comparisons were made). The go/no-go procedure requires withholding a response when the answer is thought to be “no”, whereas the go/go procedure utilised in this study calls for a different response button to be pressed for each response. Gomez, Ratcliff, and Perea (2007) conducted multiple studies comparing the two procedures, finding no differences in the core information on which decisions were based, indicating that the vital cognitive operations involved in assessing lexical status do not differ. As withholding a response may be more difficult for children, the go/go procedure was consequently selected for this study despite the non-word responses not being of any theoretical interest.

No feedback was given regarding the correctness of participant responses. The process was repeated for each narrative until all were completed, with the order of trials randomised for each participant. Comprehension questions were asked after 18 of the 60 trials, to increase the likelihood that participants were attending to the narratives. Responses to comprehension questions were recorded by the researcher, whereas RT and accuracy of participant responses to targets were recorded by the application as measures of inferencing ability.

To minimise fatigue and help maintain alertness throughout the task, five rest breaks were offered via a message on the screen every 10 trials. Participants were encouraged to use this time to leave the computer briefly to reduce eye muscle and cognitive fatigue, and experimenters interacted with each participant to maintain rapport. When participants returned from breaks, they pressed any key on the response box to continue when ready. This measure took approximately 40 minutes to complete. Upon completion of the study, participants were provided with a debriefing sheet outlining the general aims of the tasks and study provided (see Appendix E).

Results

Two participants did not meet the NARA-III accuracy score requirements detailed previously, and were consequently excluded.

Data Screening

Three conditions of RT and accuracy data were examined in the inferencing measure: bridging, predictive, and neutral. Trial RTs of more than double the mean for that condition were considered likely to reflect a lapse in focus, and consequently were coded as errors along with incorrect responses. To increase the reliability of the data set, participants who made errors on more than half of the trials in a single condition were excluded from analyses, as were participants who made errors on more than half of the narrative comprehension questions.

Additionally, although data from the non-word filler condition is not included in the statistical analysis, removal of participants on the basis of scores in this condition was conducted to reduce the possibility of participants guessing whether target words were real or not. As the filler condition consisted entirely of non-words, even if participants achieved a perfect score on the other three conditions (consisting solely of real words), it is possible that they were simply responding "yes" unless they knew for certain that the target word was not a real word. A response bias such as this could increase the possibility that some of the correct answers in the real word conditions were simply guesses, and would be reflected in high error rates for non-word targets despite low error rates for word targets. Participants with non-word accuracy rates of below 50% were therefore excluded from analyses. Lastly, removal of participants on the basis of low inference sentence comprehension scores was also carried out to increase the likelihood that participants were following instructions to engage with the narratives, rather than simply skimming the text and not attempting to retain the information they had read. These screening procedures resulted in data from an additional four participants being excluded, bringing the total number of participants included in the analysis to 30.

Remaining data were screened for outliers using a criterion of ± 3 standard deviations from the sample mean, for each participant's mean RT and overall accuracy score, for each condition. No outliers were detected for RT data, however three (negative) outliers were detected for accuracy data, two in the neutral condition and one in bridging. These outliers were winsorised, a process whereby outlier scores are changed to be the exact value of +3SD for positive outliers, and -3SD for negative, such that they are still the most extreme scores in the data set, yet still fall within our variance range limits (Lance, 2011).

Data Analysis

Comprehension Analyses. All statistical analyses were conducted at an alpha level of .05. First, a bivariate correlation was performed, to determine the relationship between comprehension scores obtained when testing on the DARC (M = 84.333, SD = 7.589) versus the NARA-III (M = 47.575, SD = 14.336). As expected, a significant positive correlation was obtained $r(28) = .388$, $p = .034$, $R^2 = .151$, (two-tailed), indicating moderately related, but distinct measures. Analyses were therefore conducted utilising scores from each measure, for comparison purposes.

Next, standardisation of scores was completed, in order to ascertain where participants from this sample were situated, relative to Australian population norms for this age group. As normative data were only available for the NARA-III, the comprehension scores from this measure were the only ones able to be standardised.

As shown on Figure 1, over 60% of the sample fell in the “average” national profile level.

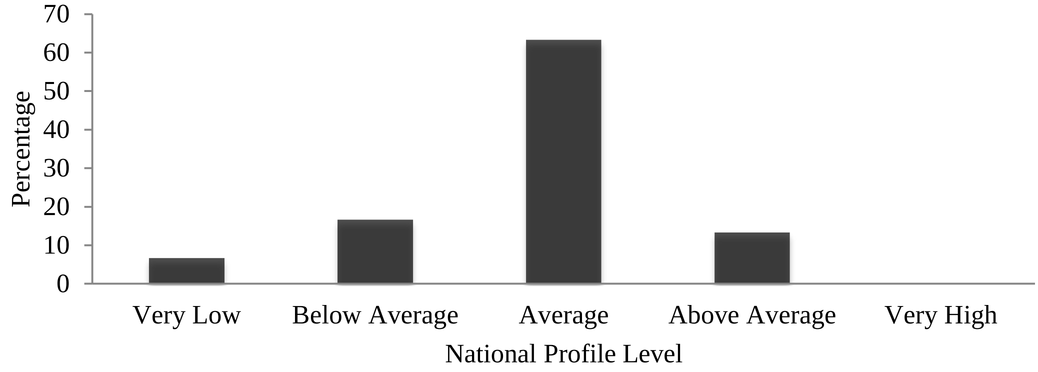

Figure 1. Percentage of participants in each national profile level rank, based on NARA-III comprehension scores.

To investigate any effects of individual differences in comprehension ability on the dependent variables, the continuous scores from each comprehension measure were transformed into median splits, categorising half of the participants into the low comprehension group, and the other half into the high comprehension group. Data were then analysed separately. Descriptive statistics are provided in Table 2.

Table 2

Descriptive Statistics for Comprehension Groups After a Median Split

	Total Group	Low Comprehension Group		High Comprehension Group	
Comprehension Measure	Median	Min	Max	Min	Max
DARC	85.000	70.000	83.333	86.667	100.000
NARA-III	50.000	22.727	47.727	52.273	75.000

Note. n = 30 in the total group, and n = 15 in the low and high comprehension groups

Reaction Time Analyses. Visual inspection of histograms, Q-Q plots, boxplots, as well as skewness, kurtosis and the Shapiro-Wilk statistic all indicated that the assumption of normality was not violated. Reaction time data was analysed using a 3x2 split-plot analysis of variance (SPANOVA), on the within-group IV of trial type (predictive inference, bridging inference, neutral) and the between-groups IV of comprehension group (high, low). This analysis was first carried out on data where the comprehension groups were based on the DARC scores. These data are presented in Figure 2.

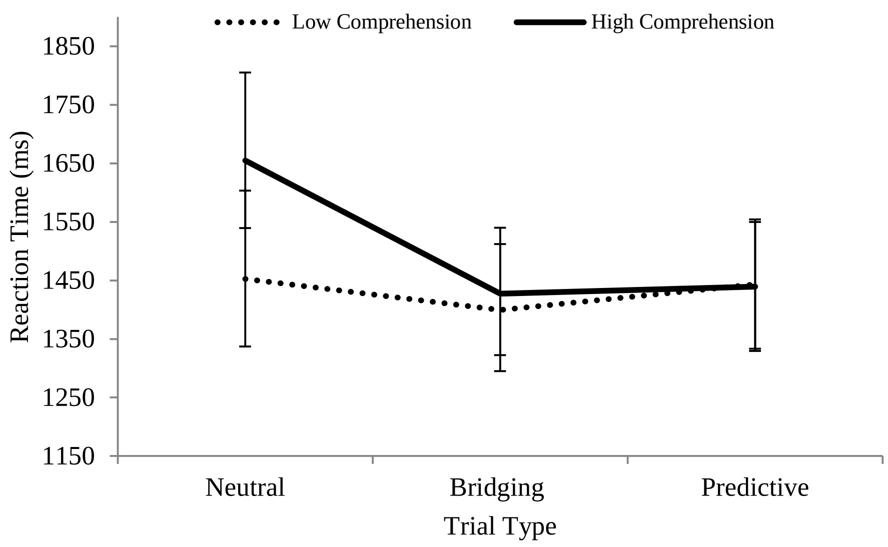

Figure 2. Mean reaction times for each trial type, as a function of comprehension ability, categorised as low comprehension (n = 15) and high comprehension (n = 15) by DARC scores. Error bars represent ±1 standard errors.

Box's M indicated no violation of the homogeneity of variance-covariance throughout the RT raw data analyses, however Mauchley's test demonstrated that the assumption of sphericity was violated, therefore Huynh-Feldt epsilon adjusted scores were reported.

No significant main effect was found for comprehension group, $F(1, 28) = .245$, $p = .624$, $\eta^2 = .009$, nor for trial type $F(1.816, 28) = 2.873$, $p = .071$, $\eta^2 = .088$.

The interaction between trial type and comprehension group was also not significant, $F(1.816, 28) = 1.605$, $p = .212$, $\eta^2 = .049$.

The 3x2 SPANOVA was repeated using the within-group IV of condition type and the between-groups IV of comprehension group, this time with “high” and “low” groups allocated according to the NARA-III scores. These data are presented in Figure 3.

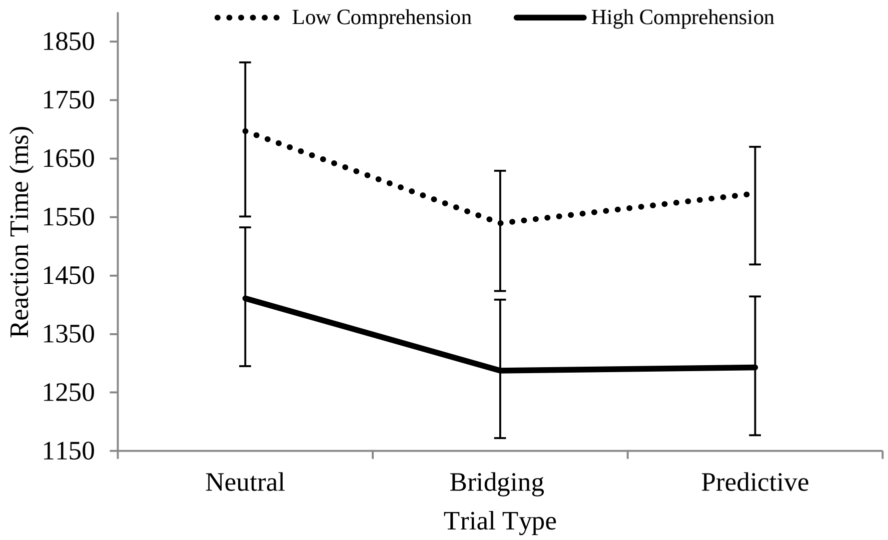

Figure 3. Mean reaction times for each trial type, as a function of comprehension ability, categorised as low comprehension ($n = 15$) and high comprehension ($n = 15$) by NARA-III scores. Error bars represent ±1 standard errors.

No significant main effect was found for comprehension group $F(1, 28) = 3.781$, $p = .062$, $\eta^2 = .119$, nor for trial type $F(1.788, 28) = 2.724$, $p = .081$, $\eta^2 = .088$. The interaction between trial type and comprehension group was also not significant, $F(1.788, 28) = .069$, $p = .916$, $\eta^2 = .002$.

As shown in both Figure 2 and 3, the RTs for low and high comprehenders in the neutral condition were dissimilar, although the lack of interaction effects indicates that this did not reach statistical significance. This suggests that there may be some existing group differences in either response speed, or the ability to make a lexical decision on target words that do not reflect inferences. Consequently, facilitation scores were calculated in order to evaluate any group differences that might exist when these existing RT differences are controlled for.

Facilitation Effect (RT) Analyses. Facilitation effect scores (ms) were calculated for each participant by subtracting the mean RT of each inference condition from the mean RT of the neutral condition in the inferencing task. These data are presented in Table 3, for comprehension groups split according to the DARC scores.

Table 3

Mean RT facilitation effects (ms) as a function of DARC Comprehension Grouping

	Low (*n* = 15)		High (*n* = 15)	
Condition	*M*	*SD*	*M*	*SD*
Bridging Facilitation	53.027	247.830	227.716	507.457
Predictive Facilitation	9.139	182.813	215.552	115.767

Following confirmation that all statistical assumptions were met, a 2x2 SPANOVA was conducted on this facilitation data, for the within-groups IV of inferencing condition (bridging, predictive) and the between-groups IV of comprehension group (low, high).

No significant main effect was found for comprehension group $F(1, 28) = 2.248$, $p = .145$, $\eta^2 = .074$, nor for inferencing condition $F(1, 28) = .338$, $p = .566$, $\eta^2 = .012$. The interaction between inferencing condition and comprehension group was also not significant, $F(1, 28) =.108$, $p = .745$, $\eta^2 = .004$.

An additional 2x2 SPANOVA was conducted on the same IVs, where comprehension group was split according to the NARA-III scores. These data are presented in Table 4.

Table 4

Mean RT facilitation effects (ms) as a function of NARA-III Comprehension Grouping

Condition	Low (n = 15)		High (n = 15)	
	M	*SD*	*M*	*SD*
Bridging Facilitation	157.557	477.265	123.187	326.902
Predictive Facilitation	106.864	395.697	118.008	317.164

In line with DARC facilitation effect analyses, no significant main effect was found for comprehension group, $F(1, 28) = .008$, $p = .931$, $\eta^2 = .001$, or for inferencing condition $F(1, 28) = .340$, $p = .565$, $\eta^2 = .012$. Additionally, the interaction between the comprehension group and inferencing condition was not significant, $F(1, 28) = .225$, $p = .639$, $\eta^2 = .008$.

Accuracy Analyses. Visual inspection of histograms, Q-Q plots, boxplots, as well as skewness, kurtosis and Shapiro-Wilk statistics all indicated that the assumption of normality was severely violated for inference accuracy data, even after the winsorisation process was conducted to reduce the effect of outliers. A ceiling effect existed across all trial types, with all participants achieving scores of 80% and above. For this reason, neither significant, nor meaningful effects from this data were considered likely. However, non-parametric testing was performed for confirmation purposes, as these analyses directly pertain to the hypotheses.

Friedman's ANOVA testing was conducted to investigate differences in accuracy by trial type (neutral, bridging, predictive). No significant effects were found, $\chi^2(2) = 2.545$, $p = .362$. Z scores from Wilcoxon tests were used to calculate effect sizes, which were very low between bridging and predictive, $r =$ □0.146, bridging and control, $r = -0.018$, and predictive and control, $r = -0.125$. Descriptive statistics are provided in Table 5.

Table 5

Descriptive Statistics for Accuracy by Trial Type

Condition	*M*	*SD*
Bridging	98.667	3.457
Predictive	97.952	4.745
Neutral	97.908	5.551

Kruskal-Wallis testing was then used to investigate differences in accuracy for each trial type (neutral, bridging, predictive), by DARC comprehension group (low and high). No significant effects were found for predictive accuracy, $H(1) = .000$, $p = 1.000$, $r = 0$, bridging accuracy, $H(1) = .910$, $p = .490$, $r = -0.123$, or neutral accuracy, $H(1) = .369$, $p = .531$, $r = -0.078$. Table 6 shows mean rank scores.

Table 6

Accuracy as a function of DARC Comprehension Grouping

	Low (n = 15)	High (n = 15)
Condition	Mean Rank	Mean Rank
Bridging	16.570	14.430
Predictive	15.500	15.500
Neutral	16.130	14.870

Kruskal-Wallis testing was then repeated for NARA-III generated comprehension groups, with no significant effects found for predictive accuracy, $H(1) = 4.462$, $p = .100$, $r = -0.273$, bridging accuracy, $H(1) = 3.308$, $p = .143$, $r = -0.235$, or neutral accuracy, $H(1) = 2.260$, $p = .215$, $r = -0.194$. Table 7 shows mean rank scores.

Table 7

Accuracy as a function of NARA-III Comprehension Grouping

	Low (n = 15)	High (n = 15)
Condition	Mean Rank	Mean Rank
Bridging	13.470	17.530
Predictive	13.500	17.500
Neutral	13.930	17.070

All participants having very high accuracy levels suggested that a speed-accuracy trade-off may exist. Two-tailed bivariate correlations were therefore performed to assess the presence of a positive correlation, specifically, whether an increase in accuracy was correlated with an increase in RT. Non-significant and small negative correlations were observed for both the predictive condition, r_s = -.170, p = .369, and the neutral condition, r_s = -.096, p = .612, , R^2 = .009. A non-significant moderate correlation was observed for the bridging condition, r_s = -.353, p = .056. These results do not indicate that a speed-accuracy trade-off effect existed.

Additional Analyses. As significant results were not found using a median split to create comprehension groups with either the DARC or the NARA-III, an exploratory analysis using tertile splits was conducted, in order to attempt to better differentiate between the two groups (i.e., high and low comprehension groups were created using the upper and lower tertile groups, respectively, with the middle tertile removed). 3x2 SPANOVAs on the same IVs (i.e., within-group IV of trial type (predictive inference, bridging inference, neutral) and the between-groups IV of comprehension group (high, low)) were again conducted for splits derived from DARC scores and NARA-III scores. Again, no significant main effects or interactions were uncovered, therefore the median split was utilised to maintain greater statistical power, due to the larger sample size in each group.

Discussion

The aim of this study was to investigate whether children with high comprehension ability differed in inference-making ability to those with low comprehension ability, and if differences exist when comprehension is categorised according to measures designed to assess low (NARA-III) versus high (DARC) level reading skills. Based on the expectation that facilitation effects would occur for target words designed to elicit inference generation, it was anticipated that participants would respond faster and more accurately to target words in the experimental inference conditions than the neutral condition. It was firstly hypothesised that the high comprehension group would have greater accuracy and faster RTs for constructing both predictive and bridging inferences than the low comprehension group, based on DARC scores. Secondly, it was hypothesised that when comprehension groups were generated using NARA-III scores, the high comprehension group would have greater accuracy and faster reaction times than the low comprehension group, just for the construction of predictive inferences. Neither hypothesis was supported.

Results Interpretation

Overall, results indicated that mean RTs did not vary significantly between the bridging and predictive conditions, nor between the neutral and inferencing conditions, with the latter providing a preliminary indication of a lack of inference-making. This was indeed further supported by the lack of significant differences observed in subsequent facilitation analyses. These results suggest that either participants were not generating inferences, or that the experimental inferencing measure did not successfully detect inference-making.

The lack of facilitation effects may have occurred as a by-product of participants focusing predominantly on the lexical decision task, (i.e., taking longer to respond in order to make an accurate decision). This was supported by the overall very high accuracy levels for all participants, resulting in a lack of significant differences between the neutral, bridging and predictive conditions. Results of the correlation conducted to assess the presence of a speed accuracy trade-off were not significant. However, with so little variation in the accuracy scores, it is likely that participants may indeed have sacrificed RT in order to maintain very high accuracy.

Moreover, it is also possible that participants may not have been making the appropriate connections between the components of a single trial. That is, participants may have focused on reading the narratives (as confirmed by correct responses to comprehension questions), then temporarily discarded this information from working memory to focus their attention on completing the lexical decision task. Consequently, rather than relating the target word to the narrative immediately prior as part of the same task, participants may have focused on the parts of each trial as separate tasks. Yuill and Oakhill (1991) found that low and high comprehenders with age-appropriate reading skills do not differ on accuracy, speed, or automaticity of single-word decoding, and that differences in ability were at the text, rather than word, level. As segmentation of the task components would result in participants simply decoding single targets, Yuill and Oakhill's (1991) findings could explain the lack of significant differences overall for RT, facilitation and accuracy.

Neither the largely experimental DARC or the standardised, extensively tested NARA-III revealed significant differences in comprehension ability between groups. Examination of grouping revealed that two thirds of participants were categorised identically by both measures, further demonstrating their related, yet distinct nature. These results provide preliminary support for the lack of meaningful differentiation between comprehension groups being relevant to the sample (i.e., indicating that it is likely that the comprehension measures utilised do not best explain the absence of evidence supporting the hypotheses). To investigate this further, NARA-III comprehension scores were categorised by standardised national profile level ranks, in order to compare the sample distribution relative to Australian population norms. This categorisation revealed that sample scores clustered around the lower end of comprehension ability, further supporting a lack of meaningful differentiation between high and low groups.

Following on from the lack of significant group differences, regardless of whether comprehension was split into low and high groups according to performance on the DARC or the NARA-III, no significant differences were observed between groups for either accuracy or RT (for predictive, bridging, or neutral conditions); nor for facilitation (for bridging or predictive). If the results of this study reflected a true lack of comprehension group differences (i.e., meaningful group distinction), this would suggest that within this age range, comprehension ability does not significantly influence either bridging or predictive inferencing ability.

However, it is worth noting that high comprehenders have consistently been found to have higher levels of familiarity, and consequently, faster RTs to targets than low comprehenders (e.g. Feagans & Appelbaum, 1986; Perfetti, 2007). Hence, even if inferences were not being generated, it would still be anticipated that high comprehenders would be reacting significantly faster than low comprehenders overall. This did not occur with the current sample, further demonstrating a lack of meaningful group differentiation. In conjunction with the very high accuracy scores for all participants, this also reinforces the likelihood of response strategies being utilised.

As high comprehenders appear to have been underrepresented in the current sample, Cain and Oakhill's (1999) finding that low comprehenders often approach reading with different goals than high comprehenders (i.e., focusing more on word reading accuracy than comprehension) could potentially explain the overall high accuracy, and the lack of facilitation effects detected.

Limitations and Future Research

Neale Analysis of Reading Ability. The NARA-III manual (Neale, 1999) dictates that testing stops after a prescribed number of reading errors are made per story. Because of this stipulation, it is possible that poor readers may have been getting the majority of comprehension questions correct, but due to reading errors, testing was ceased. Given the opportunity, these poor readers may have continued to get comprehension questions correct, in spite of poor reading accuracy. Conversely, good readers may have been answering less comprehension questions correctly, but because they were able to continue further into the test, had more opportunities to increase their comprehension score. This is of importance because comprehension is scored as the number of correct questions out of the total number of comprehension questions in the measure, irrespective of how many questions were attempted. Poor readers are therefore more likely to also have poor comprehension scores, despite what additional testing may reveal their comprehension ability to be, as detailed by Spooner et al. (2004). Hence, this could have contributed in part to the suspected overrepresentation of low comprehenders in the current study. Normative data based off scoring comprehension as a percentage of attempted questions correct may therefore be beneficial for future research, for comparison purposes.

Diagnostic Assessment of Reading Comprehension. The DARC, however, has no such reading ability restrictions on possible comprehension score. Because the DARC is purported to be measuring the underlying skills of inferencing while differentiating these from word recognition skills (August et al., 2002), it stands to reason that high and low groups should therefore have greater differences in inferencing ability when categorised by the DARC, as opposed to the NARA-III.

However, as significant differences were not found between groups based off DARC scores either, this was not evident.

A possible explanation for this lack of group distinction is that total scores were used to categorise participants by the DARC. The measure consists of four individual components: text inferencing (making inferences based on information provided in the text), knowledge access (recalling new text information from memory), text memory (accessing prior knowledge from long-term memory), and knowledge integration (integrating prior knowledge with new text information). The first two align best with skills pertaining to bridging inference-making, and the remaining two to predictive. It is possible that better differentiation between groups would occur if these component scores were used independently for assessing comprehension ability for each inference type. At present, however, the components have only low to moderate reliability when separated, and are of unequal sizes. This makes the components difficult to compare, and unlikely to produce meaningful data, hence total scores were used for the current study.

The initial purpose of the DARC was to assess reading comprehension in bilinguals learning English as a second language, who struggle with lower-level reading comprehension assessments due to issues such as lack of familiarity with vocabulary (August et al., 2006). This study is one of the first being conducted on participants who speak English as their first language, and it is possible that the DARC may not yet be appropriate for this use. Also, as suggested by August et al. (2006), the DARC may not generalise to a more heterogeneous sample of readers. Accordingly, DARC scores for this sample clustered at the high end, with a smaller range than NARA-III scores. Further development of the DARC, including equalising the number of questions in each component and further investigation of its generalisability to other samples, would likely make this measure more fit for purpose in similar future research.

Comprehension Score Distribution. As noted earlier, past research has shown SES to be positively correlated with comprehension ability. While participants were recruited from schools ranging in SES to attempt to obtain a wider range of comprehension abilities, the sample SES range was limited to the schools able to participate in this research. The analysed sample consisted predominantly of students from below average SES areas, and in line with past research findings, high comprehenders were underrepresented. Future research could therefore benefit from a more even range of SES, to aid with obtaining a more even representation of comprehension ability levels.

Most importantly, however, the sample size for the current study was likely insufficient. Time permitting, 90 participants would have been recruited, and tertile splits performed. The middle 30 participants would have been excluded from the analysis, creating a finer distinction between groups. While exploration of data from the current sample using tertile splits did not yield significant results, this was likely due to the resulting (very small) sample size. A larger sample size would not only increase the likelihood of the representation of all comprehension levels; it would also aid in producing adequate statistical power for the detection of anticipated effects, as evidenced by the small-medium effect sizes throughout these analyses. The possibility that comprehension group differences would exist in inferencing ability for this age range therefore cannot be completely ruled out. It is still possible that having more participants for the current study may still not have yielded significant differences though, due to suspected limitations of the inferencing measure.

Inferencing Measure. The inferencing task was an experimental measure, with stimuli adapted from studies by Fincher-Kiefer (1993, 1995), who consistently discovered significant facilitation effects (indicating inference generation) when the measure was used by adults. It is possible that the lack of significant differences between trial conditions in the current study pertains to the developmental trajectory of the cognitive skills and processes related to bridging and predictive inferencing. Past research has shown that younger children have the necessary knowledge to make the required inferences, however the knowledge is not related to the text in order to complete the missing details. Thus, younger children are able to construct the same inferences as older children and adults, but are less likely to do so spontaneously (e.g. Barnes et al., 1996; Cain & Oakhill, 1999). Therefore, while these skills and processes were easily able to be detected and distinguished between adults (Fincher-Kiefer, 1993, 1995), it is possible that inference-making skills and processes may not yet be sufficiently developed for children of this age group to consistently generate inferences spontaneously when using this measure. Further testing of the adapted inferencing task would help to confirm whether it is actually age-appropriate for this sample, whereby trials are individually assessed for suitability. Pilot participants of this age group could be asked to read the narrative, and then be prompted by the researcher with an inference-related question, to determine whether individual trials consistently enable participants of this age to generate the associated inference target or not.

Further, while exploring potential causes of the increased variance in the DARC neutral condition comparative to the inferencing conditions, it was discovered that some of the targets in the neutral trials may not have been truly neutral in terms of inference generation. For instance, the neutral condition narrative

"Mary was meeting a friend for lunch. After she sat down, the waiter brought over a menu. There was nothing on the menu that Mary liked so she stood up to leave" had the target "eat". While it is unlikely participants inferred that Mary did eat, it is possible that participants may have inferred that Mary did not eat. Therefore the target "eat" may still be reflective of an inference, rather than just being generally related to the scenario. As there were only ten trials per condition, the possibility of even a few trials not being truly inference neutral has the ability to impact results; therefore this could indeed account for the higher level of variance observed. Importantly, this could potentially have caused a decrease in RT difference between the neutral and inferencing conditions, therefore making it less likely that facilitation effects were able to be detected. This indicates that the aforementioned pilot testing should also be conducted on the neutral trials, to increase confidence of their inference neutrality.

Lastly, to address both task segmentation and speed accuracy trade-off issues, future research may benefit from altering the measure such that RT and accuracy are monitored in real-time, with feedback appearing if participants are answering too quickly, slowly, or incorrectly.

Conclusion

The aim of this study was to aid with the early detection of children's higher-level reading comprehension difficulties by furthering the knowledge-base on inference-making; a crucial skill underlying reading comprehension. Insight into the effect of reading comprehension level on inferencing ability could ultimately aid in more targeted and effective teaching and learning strategies. Accordingly, this study investigated the difference between high and low comprehenders in the construction of bridging and predictive inferences, and is one of the first to assess whether differences exist when comprehension is categorised according to measures designed to assess low versus high-level reading skills.

Overall, the data does not reflect meaningful categorisation between groups, irrespective of whether they were categorised according to the NARA-III or the DARC. Nor does it reflect whether inferences were generated or not, therefore interpretation of the extent of inference generation between high and low comprehenders is difficult. If these results are accurate, this would suggest that within this age range, comprehension ability does not significantly influence either bridging or predictive inferencing ability irrespective of whether comprehension is measured as a function of lower or higher-level skills. More likely, however, response strategies such as segmentation of task components or a speed accuracy trade-off, combined with an inadequate sample size to detect effects, are plausible explanations for these findings.

Limitations with NARA-III and DARC scoring may also have contributed to the lack of significant results in the current study. Further, this is the first research conducted on children who speak English as a first language using the DARC, and the first use of the adapted inferencing measure on children. The results provide evidence that levelling out the DARC subscales to increase subscale reliability would be beneficial for future research. Attempting to minimise inferencing task response strategies would also be advantageous. This could be done by altering the measure to monitor RT and accuracy in real-time, and provide feedback if participants answer too quickly, slowly, or incorrectly. Additionally, testing the adapted inferencing trials to ensure their construct validity would be beneficial. Further pilot testing of the DARC and inferencing measure is required before they can be used with confidence with this age range, in order to advance the knowledge-base. Future testing will also require a larger sample size, in order to produce adequate statistical power to observe meaningful effects. Despite its limitations, this study clearly contributes much to the literature. Importantly, it provides a preliminary foundation for future research by illuminating methodological issues that need to be addressed, in order for further investigation into this important area of research to occur.

References

August, D., Francis, D. J., & Calderón, M. (2002). *Diagnostic assessment of reading comprehension (DARC).* Washington, DC: Center for Applied Linguistics.

August, D., Francis, D. J., Hsu, H. A., & Snow, C. E. (2006). Assessing reading comprehension in bilinguals. *The Elementary School Journal*, *107*(2), 221–238. doi:10.1086/510656

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

Beeman, M. J., Bowden, E. M., & Gernsbacher, M. A. (2000). Right and left hemisphere cooperation for drawing predictive and coherence inferences during normal story comprehension. *Brain and Language*, *71*(2), 310–336. doi:10.1006/brln.1999.2268

Cain, K. (2009). Making sense of text: Skills that support text comprehension and its development. *Perspectives on Language and Literacy*, *35*(2), 11–14.

Cain, K., & Oakhill, J. (2007). Issues of causality in children's reading comprehension. In *Reading comprehension strategies*. Lawrence Erlbaum Associates.

Cain, K., & Oakhill, J. (2008). *Children's comprehension problems in oral and written language: A cognitive perspective*. Guilford Press.

Cain, K., & Oakhill, J. V. (1999). Inference making ability and its relation to comprehension failure in young children. *Reading and Writing*, *11*(5-6), 489–503. doi:10.1023/A:1008084120205

Cain, K., Oakhill, J. V., Barnes, M. A., & Bryant, P. E. (2001). Comprehension skill, inference-making ability, and their relation to knowledge. *Memory & Cognition, 29*(6), 850–859. doi:10.3758/BF03196414

Currie, J. (2008). *Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development*. National Bureau of Economic Research.

De Jong, P. F., & van der Leij, A. (2002). Effects of phonological abilities and linguistic comprehension on the development of reading. *Scientific Studies of Reading, 6*(1), 51–77. doi:10.1207/S1532799XSSR0601_03

Eternity+ Empirisoft Corporation. (2014). Retrieved from http://www.empirisoft.com/DirectIN.aspx

Feagans, L., & Appelbaum, M. I. (1986). Validation of language subtypes in learning disabled children. *Journal of Educational Psychology, 78*(5), 358. doi:10.1037/0022-0663.78.5.358

Fincher-Kiefer, R. (1993). The role of predictive inferences in situation model construction. *Discourse Processes*, *16*(1-2), 99–124.

Fincher-Kiefer, R. (1995). Relative inhibition following the encoding of bridging and predictive inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 981–995. doi:10.1037/0278-7393.21.4.981

Fincher-Kiefer, R., & D'Agostino, P. R. (2004). The role of visuospatial resources in generating predictive and bridging inferences. *Discourse Processes*, *37*(3), 205–224. doi:10.1207/s15326950dp3703_2

Flouri, E., & Buchanan, A. (2004). Early father's and mother's involvement and child's later educational outcomes. *British Journal of Educational Psychology*, *74*(2), 141–153. doi:10.1348/000709904773839806

Francis, D. J., Snow, C. E., August, D., Carlson, C. D., Miller, J., & Iglesias, A. (2006). Measures of reading comprehension: A latent variable analysis of the diagnostic assessment of reading comprehension. *Scientific Studies of Reading*, *10*(3), 301–322. doi:10.1207/s1532799xssr1003_6

Gomez, P., Ratcliff, R., & Perea, M. (2007). A model of the go/no-go task. *Journal of Experimental Psychology: General*, *136*(3), 389. doi:10.1037/0096-3445.136.3.389

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, *7*(1), 6–10. doi:10.1177/074193258600700104

Gould, J. K. (2008). *The processing of predictive and bridging inferences, with reference to age- and task-related differences* (Ph.D.). Boston University, United States -- Massachusetts.

Graesser, A. C., & Bertus, E. L. (1998). The construction of causal inferences while reading expository texts on science and technology. *Scientific Studies of Reading*, *2*(3), 247–269.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*(3), 371. doi:10.1037/0033-295X.101.3.371

Hannon, B., & Daneman, M. (2001). A new tool for measuring and understanding individual differences in the component processes of reading comprehension. *Journal of Educational Psychology*, *93*(1), 103–128. doi:10.1037/0022-0663.93.1.103

Kintsch, W. (1988). The role of knowledge in discourse comprehension: a construction-integration model. *Psychological Review*, *95*(2), 163. doi:10.1037/0033-295X.95.2.163

Lance, M. (2011). Approximate vs. monte carlo critical values for the winsorized t-test.

Landi, N. (2010). An examination of the relationship between reading comprehension, higher-level and lower-level reading sub-skills in adults. *Reading and Writing*, *23*(6), 701–717. doi:10.1007/s11145-009-9180-z

McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review*, *99*(3), 440. doi:10.1037/0033-295X.99.3.440

McNamara, D. S. (2007). *Reading comprehension strategies*. Lawrence Erlbaum Associates.

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*(1), 199–211. doi:10.1044/1092-4388(2004/017)

Nation, K., & Snowling, M. J. (1998). Semantic processing and the development of word-recognition skills: Evidence from children with reading comprehension difficulties. *Journal of Memory and Language*, *39*(1), 85–101. doi:10.1006/jmla.1998.2564

Neale, M. D. (1999). *Neale analysis of reading ability: Manual - Australian Standardisation* (third ed.). Melbourne, Australia: ACER Press, Australian Council for Educational Research Limited.

Oakhill, J., & Cain, K. (2000). Children's difficulties in text comprehension: Assessing causal issues. *Journal of Deaf Studies and Deaf Education*, *5*(1), 51–59. doi:10.1093/deafed/5.1.51

Oakhill, J. V., Cain, K., & Bryant, P. E. (2003). The dissociation of word reading and text comprehension: Evidence from component skills. *Language and Cognitive Processes*, *18*(4), 443–468. doi:10.1080/01690960344000008

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, *11*(4), 357–383. doi:10.1080/10888430701530730

Perfetti, C. A. (1985). *Reading ability*. Oxford University Press. Retrieved from http://doi.apa.org/psycinfo/1985-97290-000

Perfetti, C. A., & Hogaboam, T. (1975). Relationship between single word decoding and reading comprehension skill. *Journal of Educational Psychology*, *67*(4), 461. doi:10.1037/h0077013

Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. doi:10.1002/9780470757642.ch13

Radvansky, G. A., & Dijkstra, K. (2007). Aging and situation model processing. *Psychonomic Bulletin & Review*, *14*(6), 1027–1042. doi:10.3758/BF03193088

Schmalhofer, F., McDaniel, M. A., & Keefe, D. (2002). A unified model for predictive and bridging inferences. *Discourse Processes*, *33*(2), 105–132. doi: 10.1207/S15326950DP3302_01

Shankweiler, D. (1989). How problems of comprehension are related to difficulties in decoding. *Phonology and Reading Disability: Solving the Reading Puzzle*, 35–68.

Shankweiler, D., Lundquist, E., Katz, L., Stuebing, K. K., Fletcher, J. M., Brady, S., … Shaywitz, B. A. (1999). Comprehension and decoding: Patterns of association in children with reading difficulties. *Scientific Studies of Reading, 3*(1), 69–94. doi:10.1207/s1532799xssr0301_4

Spooner, A. L., Baddeley, A. D., & Gathercole, S. E. (2004). Can reading accuracy and comprehension be separated in the Neale analysis of reading ability? *British Journal of Educational Psychology*, *74*(2), 187–204. doi:10.1348/000709904773839833

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. Academic Press New York.

White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin*, *91*(3), 461. doi:10.1037/0033-2909.91.3.461

Yuill, N., & Oakhill, J. (1991). *Children's problems in text comprehension: An experimental investigation*. Cambridge University Press.

Zwaan, R. A. (1999). Situation models: The mental leap into imagined worlds. *Current Directions in Psychological Science*, *8*(1), 15–18. doi:10.1111/1467-8721.00004

Zwaan, R. A. (2004). The immersed experiencer: Toward an embodied theory of language comprehension. *Psychology of Learning and Motivation*, *44*, 35–62.

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science, 13*(2), 168–171.

Instructions to Authors

Journal of Educational Psychology

Prior to submission, please carefully read and follow the submission guidelines detailed below. Manuscripts that do not conform to the submission guidelines may be returned without review.

Submission
Submit manuscripts electronically via the Manuscript Submission Portal.

Steve Graham, EdD
Arizona State University

General correspondence may be directed to the Editor's Office.

In addition to addresses and phone numbers, please supply email addresses, as most communications will be by email. Fax numbers, if available, should also be provided for potential use by the editorial office and later by the production office.

Masked Review Policy
The *Journal* has a masked review policy, which means that the identities of both authors and reviewers are masked. Authors should make every effort to see that the manuscript itself contains no clues to their identities.

Authors should never use first person (I, my, we, our) when referring to a study conducted by the author(s) or when doing so reveals the authors' identities, e.g., "in our previous work, Johnson et al., 1998 reported that…" Instead, references to the authors' work should be in third person, e.g., "Johnson et al. (1998) reported that…."

The authors' institutional affiliations should also be masked in the manuscript.

Include the title of the manuscript along with all authors' names and institutional affiliations in the cover letter. The first page of the manuscript should omit the authors' names and affiliations, but should include the title of the manuscript and the date it is submitted.

Responsibility for masking the manuscript rests with the authors; manuscripts will be returned to the author if not appropriately masked. If the manuscript is accepted, authors will be asked to make changes in wording so that the paper is no longer masked.

If your manuscript was mask reviewed, please ensure that the final version for production includes a byline and full author note for typesetting.

Manuscript Guidelines
Adequate description of participants is critical to the science and practice of educational psychology; this allows readers to assess the results, determine generalizability of findings, and make comparisons in replications, extensions, literature reviews, or secondary data analyses. Authors should see guidelines for sample/subject description in the *Publication Manual*.

Appropriate indexes of effect size or strength of relationship should be incorporated in the results section of the manuscript (see p. 34 of the *Manual*). Information that allows the reader to assess not only the significance but also the magnitude of the observed effects or relationships clarifies the importance of the findings.

Manuscript Preparation
Prepare manuscripts according to the *Publication Manual of the American Psychological Association* (6th edition). Manuscripts may be copyedited for bias-free language (see Chapter 3 of the *Publication Manual*).

Double-space all copies. Other formatting instructions, as well as instructions on preparing tables, figures, references, metrics, and abstracts, appear in the *Manual*.

Below are additional instructions regarding the preparation of display equations, computer code, and tables.

Display Equations
We strongly encourage you to use MathType (third-party software) or Equation Editor 3.0 (built into pre-2007 versions of Word) to construct your equations, rather than the equation support that is built into Word 2007 and Word 2010. Equations composed with the built-in Word 2007/Word 2010 equation support are converted to low-resolution graphics when they enter the production process and must be rekeyed by the typesetter, which may introduce errors.

To construct your equations with MathType or Equation Editor 3.0:

- Go to the Text section of the Insert tab and select Object.
- Select MathType or Equation Editor 3.0 in the drop-down menu.

If you have an equation that has already been produced using Microsoft Word 2007 or 2010 and you have access to the full version of MathType 6.5 or later, you can convert this equation to MathType by clicking on MathType Insert Equation. Copy the equation from Microsoft Word and paste it into the MathType box. Verify that your equation is correct, click File, and then click Update. Your equation has now been inserted into your Word file as a MathType Equation.

Use Equation Editor 3.0 or MathType only for equations or for formulas that cannot be produced as Word text using the Times or Symbol font.

Computer Code
Because altering computer code in any way (e.g., indents, line spacing, line breaks, page breaks) during the typesetting process could alter its meaning, we treat computer code differently from the rest of your article in our production process. To that end, we request separate files for computer code.

In Online Supplemental Material
We request that runnable source code be included as supplemental material to the article. For more information, visit Supplementing Your Article With Online Material.

In the Text of the Article
If you would like to include code in the text of your published manuscript, please submit a separate file with your code exactly as you want it to appear, using Courier New font with a type size of 8 points. We will make an image of each segment of code in your article that exceeds 40 characters in length. (Shorter snippets of code that appear in text will be typeset in Courier New and run in with the rest of the text.) If an appendix contains a mix of code and explanatory text, please submit a file that contains the entire appendix, with the code keyed in 8-point Courier New.

Tables
Use Word's Insert Table function when you create tables. Using spaces or tabs in your table will create problems when the table is typeset and may result in errors.

If your manuscript was mask reviewed, please ensure that the final version for production includes a byline and full author note for typesetting.

Review APA's Checklist for Manuscript Submission before submitting your article.

Submitting Supplemental Materials
APA can place supplemental materials online, available via the published article in the PsycARTICLES® database. Please see Supplementing Your Article With Online Material for more details.

Abstract and Keywords
All manuscripts must include an abstract containing a maximum of 250 words typed on a separate page. After the abstract, please supply up to five keywords or brief phrases.

References
List references in alphabetical order. Each listed reference should be cited in text, and each text citation should be listed in the References section.

Examples of basic reference formats:

- **Journal Article:**
  Hughes, G., Desantis, A., & Waszak, F. (2013). Mechanisms of intentional binding and sensory attenuation: The role of temporal prediction, temporal control, identity prediction, and motor prediction. *Psychological Bulletin, 139*, 133–151. http://dx.doi.org/10.1037/a0028566
- **Authored Book:**
  Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach.* Cambridge, MA: MIT Press.
- **Chapter in an Edited Book:**
  Gill, M. J., & Sypher, B. D. (2009). Workplace incivility and organizational trust. In P. Lutgen-Sandvik & B. D. Sypher (Eds.), *Destructive organizational communication: Processes, consequences, and constructive ways of organizing* (pp. 53–73). New York, NY: Taylor & Francis.

Figures
Graphics files are welcome if supplied as Tiff or EPS files. Multipanel figures (i.e., figures with parts labeled a, b, c, d, etc.) should be assembled into one file.

The minimum line weight for line art is 0.5 point for optimal printing.

For more information about acceptable resolutions, fonts, sizing, and other figure issues, please see the general guidelines.

When possible, please place symbol legends below the figure instead of to the side.

APA offers authors the option to publish their figures online in color without the costs associated with print publication of color figures.

The same caption will appear on both the online (color) and print (black and white) versions. To ensure that the figure can be understood in both formats, authors should add alternative wording (e.g., "the red (dark gray) bars represent") as needed.

For authors who prefer their figures to be published in color both in print and online, original color figures can be printed in color at the editor's and publisher's discretion provided the author agrees to pay:

- $900 for one figure
- An additional $600 for the second figure
- An additional $450 for each subsequent figure

Permissions
Authors of accepted papers must obtain and provide to the editor on final acceptance all necessary permissions to reproduce in print and electronic form any copyrighted work, including test materials (or portions thereof), photographs, and other graphic images (including those used as stimuli in experiments).

On advice of counsel, APA may decline to publish any image whose copyright status is unknown.

- Download Permissions Alert Form (PDF, 13KB)

Publication Policies
APA policy prohibits an author from submitting the same manuscript for concurrent consideration by two or more publications.

See also APA Journals® Internet Posting Guidelines.

APA requires authors to reveal any possible conflict of interest in the conduct and reporting of research (e.g., financial interests in a test or procedure, funding by pharmaceutical companies for drug research).

- Download Disclosure of Interests Form (PDF, 38KB)

Authors of accepted manuscripts are required to transfer the copyright to APA.

- For manuscripts **not** funded by the Wellcome Trust or the Research Councils UK
  Publication Rights (Copyright Transfer) Form (PDF, 83KB)
- For manuscripts funded by the Wellcome Trust or the Research Councils UK
  Wellcome Trust or Research Councils UK Publication Rights Form (PDF, 34KB)

Ethical Principles

It is a violation of APA Ethical Principles to publish "as original data, data that have been previously published" (Standard 8.13).

In addition, APA Ethical Principles specify that "after research results are published, psychologists do not withhold the data on which their conclusions are based from other competent professionals who seek to verify the substantive claims through reanalysis and who intend to use such data only for that purpose, provided that the confidentiality of the participants can be protected and unless legal rights concerning proprietary data preclude their release" (Standard 8.14).

APA expects authors to adhere to these standards. Specifically, APA expects authors to have their data available throughout the editorial review process and for at least 5 years after the date of publication.

Authors are required to state in writing that they have complied with APA ethical standards in the treatment of their sample, human or animal, or to describe the details of treatment.

- Download Certification of Compliance With APA Ethical Principles Form (PDF, 26KB)

The APA Ethics Office provides the full Ethical Principles of Psychologists and Code of Conduct electronically on its website in HTML, PDF, and Word format. You may also request a copy by emailing or calling the APA Ethics Office (202-336-5930). You may also read "Ethical Principles," December 1992, *American Psychologist*, Vol. 47, pp. 1597–1611.

Appendix A

Parents and Participants Consent Forms

Consent Form A

Visual Imagery and Reading Comprehension Level: A Situation Model Investigation

Parent

1. I agree voluntarily for my child to take part in this study.

2. I have read the Information Sheet provided and been given a full explanation of the purpose of this study, of the procedures involved and of what is expected of my child. The researcher has answered all my questions and has explained the possible problems that may arise as a result of my child's participation in this study.

3. I have discussed with my child what participation in this study involves and he/she has agreed voluntarily to participate, as indicated by his/her completion of this consent form.

4. I understand my child is free to withdraw from the study at any time without needing to give any reason.

5. I understand that I and my child will not be identified in any publication arising out of this study.

6. I understand that my child's name and identity will be stored separately from the data, and these are accessible only to the investigators. All data provided by my child will be analysed anonymously using code numbers.

7. I understand that all information provided by my child is treated as confidential and will not be released by the researcher to a third party unless required to do so by law.

8. *Optional: please tick the box if you answer yes to this question.*

☐ I would like to enter the prize draw to win a $100 Coles shopping voucher.

Signature of Parent: ________________________ Date: …..../..…../…….

Child

-I would like to do the paper and computer tasks.

-I am happy for you to ask me questions about the tasks.

-I know that I can choose not to answer your questions if I want to.

-I know that I can choose to stop doing the tasks at anytime.

Optional- please tick the box if you answer yes to this question.

□ I would like to go into the draw to win an iPod Touch

Signature of Child: ________________________ Date: ……./……./…….
(Name)

Signature of Investigator: _________________________ Date: ..…../…..../…….

Consent Form B

Visual Imagery and Reading Comprehension Level: A Situation Model Investigation

Parent

1. I do **not** consent to my child participating in this study/my child is not eligible to participate in this study, but I would like to enter the prize draw to win a $100 Coles shopping voucher.

Name of Parent: ________________________________

Child

I do not want to take part in the tasks/I am not able to take part in the tasks, but I would like to go into the draw to win an iPod Touch.

____

Name of Child

Appendix B

Diagnostic Assessment of Reading Comprehension Instructions

Note: These Materials have been removed from this electronic thesis copy.

Appendix C

Inferencing Task Instructions

1. Run the input file (.csv) on DirectRT.
2. Enter demographic data for participant

Give participant the piece of paper with example instructions.
"Today we are going to do a reading task on the computer. During the task you will read several short stories and then after each story you will see a word, your task is to read each story carefully and then indicate whether the word was a real word or not by pressing a button on this response box (point to box)."

"I will now explain the instructions, and then you can practice doing the task so you get used to pressing the buttons. After that we will go on to the full task."
"So, first you will see the words "get ready" appear for a few seconds like this": *****Point to - Get Ready***

"Then the first line of a story will appear. Each story is either 3 or 4 lines long and will appear on the screen one line at a time. After you read each line, press any of the middle buttons on the box to make the next line appear."
"Once you have finished the whole story, a cross will appear on the screen which looks like this:" *****On the inferencing session example page, point to---- +***

"This cross means you need to get ready because a word is about to appear."
"When the word appears: Press the "Yes" button if this word is a real word. Press the "No" button if this word is not a real word."
*****Now point to---*** **target word on inferencing session example page** and ask—"so what button would you press if this word appeared?"
If incorrect.. explain response keys again/explain that "throw" is a real word.
If correct…"Great, that is correct. It is important during the task that you also make your decision as quickly as possible, but also try and make sure that you press the correct button."
"After you have made your decision, you will see the "get ready" sign again and will then continue on to the next story."

It is also important that you read each story carefully and pay attention, because after you have read a few stories the researcher will ask you some questions about them"
"Before we start, do you have any questions?"

"Okay, let's do some practice trials, and then we will see if you are ready to go on to the full task. It is okay if you get some answers wrong during the practice- it is just to help you get used to pressing the right buttons."

Start Task
-ensure participant has left hand over the "yes" button, and right hand over "no" (explain that they should keep their hands like this during the task so they can respond as quickly as possible)
- if participant wants to do the practice trials again then restart task (up to two runs of practice trials only).

After Practice
-explain: There will be scheduled rest breaks during the task, but rest breaks can also be taken during the question times if needed- (but should try not to stop doing the task at other times).
-ask if participant has any other questions before starting.

Inferencing Session Example

Get Ready

The man looked up at the sky.

He could see the moon shining brightly.

He went inside to get his camera.

+

THROW

“yes” = real word

“no” = not a real word

Appendix D

Inferencing Task Trial Examples

<u>Neutral Condition:</u>

The two boys planned to meet on the playground after school.
One of the boys was upset because the other boy had made fun of his big ears.
They looked at each other and then one of them put out his hand.
Target: FIGHT

Bobby brought home his new pet and put it in the cage.
He soon realised he didn't have any toys for it to play with.
He picked up his car keys and wallet.
Target: FEED

Jane was running late for dinner as she parked her car.
When she returned to her car it was very dark outside.
She didn't see the coin on the footpath.
Target: TRIP

<u>Bridging Condition:</u>

The bride walked into the chapel and she couldn't believe that today was the big day.
She was very nervous about her wedding.
Near the end of the ceremony, the bride suddenly felt uneasy.
Luckily, the groom was there to pick her up off the floor.
Target: FAINT

David was very hungry as he sat down in front of the big bowl of spaghetti.
He enjoyed his meal and ate very quickly, getting sauce everywhere.
Some of the sauce dripped on to his shirt.
David was annoyed that his favourite shirt was now ruined.
Target: STAIN

Alan was the friendliest guy in his office.
Tonight he was having a big party for his work friends.
The party wasn't as lively as he wanted, so he put on an upbeat song.
Alan's feet were sore the next morning.
Target: DANCE

Predictive Condition:

The florist and her cat were busy at work in the garden.
The bees had formed their nest in the tulip bed.
Reaching down into the tulips the nest was disturbed.
Target: STING

The man was anxious to start his summer garden.
Vegetable seeds had been purchased.
The garden dirt was prepared and the holes were dug.
Target: PLANT

The man was very tired.
He drove home on a winding road.
At a sharp curve, he accidently turned off toward the large tree.
Target: CRASH

Filler Condition:

Jo was feeling bored as there was nothing on TV.
Her mum arrived home with her favourite cartoon to watch.
She spent the afternoon sitting happily on the floor.
Jo was sad when she had to turn the TV off and go have dinner.
Target: MAWT

The new worker was very nervous about her first meeting with the boss.
Several cups of coffee had been drunk to help her stay alert.
The boss was known for being grumpy.
Target: EMUNG

The husband complained to his wife that they were eating too many sweets.
He hoped that she would stop buying ice cream.
They were both getting too heavy and summer was coming.
Target: BRAUNE

Appendix E

Participant Debriefing Sheet

Visual Imagery and Reading Comprehension: A Situation Model Investigation

The aim of this study was to see what types of things children experience when they are reading and whether they imagine the objects they read about.

Some of the tasks you did were designed to assess how you read and understand stories. Others were designed to look at different types of visual imagery (how you imagine and “see” things in your mind).

Do you remember the reading tasks you did? This is what they were about;

- During the two tasks where you read stories aloud to the researcher, we were trying to see whether there is any difference between how you read and understand stories about things that are made up (for example, snerps!), compared to stories about things that are real (like playing in a tree house).
- In the reading task on the computer, we were trying to see whether having to do another task while reading (remembering letters or dots) makes it harder to understand and remember the story.
- The reading task with the picture and word cards was used to see whether it is harder to put story events in order if there was a flashback in the story.

These were the imagery tasks;

- The pen and paper task with the strange symbols was used to discover whether you rotated the symbols using only your imagination.
- The computer tasks with the grid and the ‘x’s or ‘o’s looked at whether you held a pattern in your mind for some time and remembered all the parts of the pattern.
- The task with the booklet full of different patterns was trying to see whether you could figure out which of the options was the missing piece.
- The computer task where you had to remember sets of numbers was seeing how many numbers people can remember using only their memory (which is why you couldn’t write the numbers down!).

There were no right or wrong answers in these tasks. Everyone has a different way of reading and imagining things!

Thank you for taking part!