RUNNING HEAD: ORTHOGRAPHIC AND SEMANTIC LEARNING IN POOR COMPREHENDERS

Investigating orthographic and semantic aspects of word learning in poor comprehenders

Jessie Ricketts, DVM Bishop and Kate Nation

University of Oxford

Correspondence:

Jessie Ricketts

Department of Experimental Psychology

University of Oxford

South Parks Road

Oxford, OX1 3UD, UK

Email: jessie.ricketts@psy.ox.ac.uk

Abstract

This study compared orthographic and semantic aspects of word learning in children who differed in reading comprehension skill. Poor comprehenders and controls matched for age (9-10 years), nonverbal ability and decoding skill were trained to pronounce 20 visually presented nonwords, 10 in a consistent way and 10 in an inconsistent way. They then had an opportunity to infer the meanings of the new words from story context. Orthographic learning was measured in three ways: the number of trials taken to learn to pronounce nonwords correctly, orthographic choice and spelling. Across all measures, consistent items were easier than inconsistent items and poor comprehenders did not differ from control children. Semantic learning was assessed on three occasions, using a nonword-picture matching task. While poor comprehenders showed equivalent semantic learning to controls immediately after exposure to nonword meaning, this knowledge was not well-retained over time. Results are discussed in terms of the language and reading skills of poor comprehenders and in relation to current models of reading development.

Investigating orthographic and semantic aspects of word learning in poor comprehenders

Poor comprehenders are children who experience difficulties with reading comprehension, despite age-appropriate reading accuracy. Many experiments have shown that poor comprehenders are poor at text-level processes such as making inferences and comprehension monitoring (for reviews see Cain & Oakhill, 2007; Nation, 2005). However, poor comprehenders' difficulties are not restricted to the comprehension of written text. They experience relative weaknesses in comprehending orally presented sentences and discourse, and with listening comprehension more generally (Catts, Adlof, & Weismer, 2006; Nation, Clarke, Marshall, & Durand, 2004). Narrative expression is also compromised in both written (Cragg & Nation, 2006) and oral (Cain, 2003) modalities.

Researchers studying poor comprehenders take care to select children whose reading accuracy is age-appropriate; nevertheless, there is some evidence that poor comprehenders show relative weaknesses when reading exception words – words that have atypical mappings between spelling and sound (e.g., *break*, *yacht*). Nation and Snowling (1998) first reported that poor comprehenders were significantly less accurate at reading exception words than skilled comprehenders, despite the two groups being matched for phonological decoding (nonword reading) and nonverbal reasoning scores. This finding has since been replicated (Ricketts, Nation, & Bishop, 2007).

Exception words can only be partially decoded using the alphabetic principle (Share, 1995). This suggests that good decoding skills alone are not enough to support efficient exception word reading. Various proposals have been made about additional skills that might underpin exception word reading. One is that orthographic processing skills are important (e.g., Harm & Seidenberg, 1999; Manis, Seidenberg, Doi, McBride-Chang, & Petersen, 1996). On this view, the exception word reading deficit seen in poor comprehenders may be due to weak orthographic knowledge. Typically, orthographic processing skills are assessed using tasks such as orthographic choice (e.g., which is a word: *assure* or *ashure*?) or by measuring proxy variables such as print exposure. It is worth noting that such measures are not independent of the word recognition process and therefore it is not clear whether such tasks measure word recognition itself, rather than being an index of a separable orthographic construct (for a fuller discussion of this issue see Castles & Nation, 2006; Burt, 2006). Notwithstanding the lack of consensus as to what constitutes "orthographic processing", in our research we have found that poor and skilled comprehenders do not differ in two measures of orthographic knowledge - print exposure and orthographic choice (Ricketts et al., 2007).

Nation and Snowling (1998) presented an alternative explanation, suggesting that poor comprehenders' exception word reading deficits may be a consequence of semantic weaknesses in the oral domain. Their theorizing was inspired by a connectionist model of visual word recognition and its development (Plaut, McClelland, Seidenberg, & Patterson, 1996). Simulations of this model revealed that a word recognition system trained with input from semantics was better able to compute correct pronunciations of exception words than a model trained without a semantic contribution (Harm & Seidenberg, 2004; Plaut et al., 1996) thus suggesting that semantic knowledge may play a direct role in the reading of exception words. Similarly, Keenan and Betjemann (2007) suggested that semantic knowledge - knowledge of word meanings - could provide compensatory support for exception word reading where mappings between spelling and sound are weak. For example, if an exception word is unknown, a child will have no choice but to attempt to decode it. This would result in a mispronunciation that is similar to the target word e.g., pronouncing *yacht* to rhyme with *matched* rather than *pot*. A child who has good semantic knowledge and a large vocabulary is more likely to be familiar with the exception word that they are trying to read. Also, Bowey and Rutherford (2007) suggested that a child with stronger vocabulary skills is more likely to vary the pronunciation of an unknown word to read the word successfully. Consistent with the idea that good vocabulary skills contribute to successful exception word reading, vocabulary predicts later exception word reading above and beyond decoding and several studies find oral vocabulary deficits in some poor comprehenders (Cain, Oakhill, & Lemmon, 2004; Nation & Snowling, 1998, 2004; Ricketts et al., 2007; but see Stothard & Hulme, 1992). Mirroring these findings, children with selective weaknesses in exception word reading also show concomitant weaknesses in oral vocabulary (Bowey & Rutherford, 2007; Byrne, Freebody, & Gates, 1992).

In this study we sought to investigate word learning in poor comprehenders. In order to build a sight vocabulary, a child needs to make links between phonological and orthographic information, and between these representations and a word's semantic characteristics. According to the Lexical Quality Hypothesis (Perfetti & Hart, 2002) a good quality lexical representation is one where stored phonological, orthographic and semantic information about a word is well integrated. This hypothesis is underspecified in that it does not provide more explicit predictions about which aspect of the representation is most important. However, it does provide a useful framework for considering visual word learning in suggesting that failure to learn a novel word might be a consequence of weak phonological, semantic or orthographic knowledge, or of an inability to develop associations between any of these representations. Poor comprehenders do not have difficulty processing phonology (Catts et al., 2006; Nation et al., 2004), with making consistent links between phonology and orthography (i.e., regular word reading, Nation & Snowling, 1998) or with associating a new phonological form to its new referent (Nation et al., 2007). Yet, we know that at least some poor comprehenders have low vocabulary knowledge and these children also struggle to read exception words, despite performing at a similar level to controls on other tasks thought to tap orthographic knowledge such as print exposure and orthographic choice (Ricketts et al., 2007).

This experiment examined the acquisition of semantic and orthographic representations for new words that have consistent versus inconsistent spelling-sound mappings. Castles and Holmes (1996) explored visual word learning in children with surface dyslexia - children who show particular difficulty reading exception words. They sought to investigate exception word learning by training children to pronounce printed nonwords such as *bouch*. The nonwords were assigned an inconsistent pronunciation that subjects would not produce on the basis of grapheme-phoneme conversion rules - *bouch* was pronounced to rhyme with *touch* rather than *couch*. Castles and Holmes found that children with surface dyslexia had difficulty with this task, suggesting that they may have exception word reading difficulties because they have difficulty learning exception words. Using a similar paradigm Bailey Manis, Pedersen and Seidenberg (2004) extended this research and found that children with dyslexia were poorer than age-matched controls at learning both consistent and inconsistent items.

Castles & Holmes (1996; also Bailey et al., 2004) examined the process of forming associations between orthography and phonology – a process we term orthographic learning. To our knowledge orthographic learning has not been investigated in poor comprehenders. However, Cain and colleagues (Cain, Oakhill, & Elbro, 2003; Cain et al., 2004) have looked at vocabulary acquisition in poor comprehenders. In their research poor and skilled comprehenders are exposed to novel words (nonwords) in a context that provides cues for their meaning to be inferred. Poor comprehenders were poor at inferring the meaning of the novel words. Nation, Snowling and Clarke (2007) also report a study showing that poor comprehenders have difficulty learning semantic information about novel objects. Thus it seems that as well as having poor existing oral vocabulary, poor comprehenders show weaker learning of semantic information for novel vocabulary – a process we refer to as semantic learning.

In this study we adapted the paradigm used by Castles and Holmes (1996) and Bailey et al. (2004) to probe semantic as well as orthographic aspects of word learning. We trained poor and skilled comprehenders to pronounce 10 nonwords in a consistent way, and 10 in an inconsistent way. Given poor comprehenders' deficit in reading exception words (Nation & Snowling, 1998, 2004; Ricketts et al., 2007), we anticipated that they might find it more difficult to learn the inconsistent nonwords, relative to control children. After training, poor and skilled comprehenders were exposed to the 20 nonwords embedded in story context (cf. Cain et al., 2003; Cain et al., 2004). Ten nonwords (five consistent and five inconsistent) were presented in an unhelpful context, which provided little information about their meaning. The remaining 10 nonwords were presented in a helpful context, which provided cues to each nonword meaning. This was done to assess the children's ability to use context to infer the meaning of new words. Based on previous studies (Cain et al., 2003; Cain et al., 2004; Nation et al., 2007), we hypothesized that poor comprehenders would learn fewer novel word meanings.

In summary, our primary aim was to investigate whether poor comprehenders have difficulty with orthographic and semantic aspects of word learning. To our knowledge this is the first time that semantic and orthographic learning have been contrasted, and the first time that orthographic learning has been investigated in poor comprehenders. We predicted that poor comprehenders would have difficulty with semantic learning as they show poor semantic knowledge (e.g., Nation & Snowling, 1998) and do not use context to learn the meaning of new words in other word learning experiments (Cain et al., 2003; Cain et al., 2004). We also predicted that like children with dyslexia, poor comprehenders might show poor orthographic learning overall, and perhaps particular difficulty learning inconsistent items (cf. Bailey et al., 2004; Castles & Holmes, 1996). However, this prediction was slightly tempered by our previous finding that poor comprehenders do not perform poorly on some orthographic tasks, despite being poor at reading exception words (Ricketts et al., 2007).

Method

Participants

An initial screening phase to select poor comprehenders and controls was conducted approximately 10 months prior to the experiment. We assessed 81children (see below for details of screening measures) aged between 8 years and 8 months and 9 years and 9 months. Children attended schools serving socially mixed catchment areas in

Oxford and Middlesex, all spoke English as a first language, and none had any recognized special educational needs. Our screening battery comprised the following tests:

Nonverbal reasoning. Nonverbal reasoning was measured using the Matrix Reasoning subtest of the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999). This subtest assesses nonverbal reasoning using a pattern completion task. The WASI provides norms for individuals aged 6-89 years.

Decoding. Decoding was assessed using the Phonemic Decoding component of the Test of Word Reading Efficiency (TOWRE; Torgesen, Wagner, & Rashotte, 1999). In this test children are asked to read a list of nonwords of increasing length and difficulty as quickly as they can. Efficiency is indexed by the number of nonwords decoded correctly in 45 seconds. The test provides norms for individuals aged 6-24 years.

Reading comprehension. Reading comprehension was assessed using the Neale Analysis of Reading Ability-II (NARA-II; Neale, 1997). In the NARA-II children read aloud passages of connected text and then answer comprehension questions relating to each passage. Some questions can be answered with reference to verbatim memory while others require inferences to be made (Bowyer-Crane & Snowling, 2005). This yields a measure of text reading accuracy and a measure of reading comprehension. The test provides norms for children aged 6-12 years.

From this sample, 15 poor comprehenders (11 female and 4 male) and 15 controls (11 female and 4 male) were selected, according to the following criteria. Poor comprehenders obtained NARA-II reading comprehension scores more than one standard deviation below the test mean (< 85) and controls' scores were well into the average range or above (> 95). Groups were matched for chronological age, nonverbal ability and decoding skill with all children performing well within the average range (or above). The top portion of Table 1 shows the mean chronological age of participants at the time of screening, summarizes performance of groups on screening measures, and provides *F* values for group comparisons (for a more detailed description of this sample see Ricketts, Nation & Bishop, 2007). Note that although matched for decoding skill, comprehension groups did differ significantly in their NARA-II reading accuracy scores. However, text reading accuracy in poor comprehenders was at an age-appropriate level and all children scored in the normal range.

Approximately 10 months later and at the same time as running our word learning experiment, we proceeded to assess their children's language and reading skills in more detail, using the following background measures:

Phoneme deletion. Phonological skills were assessed using a phoneme deletion task (see McDougall, Hulme, Ellis, & Monk, 1994 for materials). Children were presented with a nonword and were asked to tell the experimenter which word would remain if they took away a particular sound (e.g., bice → ice, stip → sip, cloof → clue). For example if presented with "bice" and asked to take away the sound /b/ they would be expected to respond "ice". Children were required to delete sounds from the beginning, middle and end of nonwords. Most phonemes to be deleted were from consonant clusters. Two practice trials were administered to ensure that children understood the demands of the task; then test items were administered in order of difficulty. The maximum score was 22 and a proportion correct score was calculated for each child.

Vocabulary. Vocabulary was assessed using the Multiple Contexts subtest of the Test of Word Knowledge (TOWK; Wiig & Secord, 1992). This subtest is a measure of expressive vocabulary in which children are presented with a set of words with multiple meanings (e.g., bat) and are required to provide two distinct definitions for each word (e.g., the thing you hit a ball with, an animal that flies). The TOWK provides norms for children aged 5-17 years.

Exception word reading. Children read 70 words presented one at a time in the middle of a computer screen using the E-Prime program (Schneider, Eschman, & Zuccolotto, 2002) to randomize presentation order. Children read each word aloud and accuracy was scored by the experimenter. Children read 30 exception words. The remaining 40 words varied in consistency and were included to distract attention from the exception items, so that children would not be immediately alerted to the fact that they were reading ‘strange’ words. A proportion correct score was calculated for each child.

The lower portion of Table 1 shows the mean chronological age of poor comprehenders and controls at this time point and summarizes group performance on background measures with *F* values for group comparisons. Consistent with previous work, poor comprehenders achieved lower scores on tests of vocabulary knowledge and exception word reading than control children[1], but the two groups did not differ in phonological skill.

Insert Table 1 about here

Experimental materials and procedure

Over two days, children were trained to pronounce nonwords in either a consistent or inconsistent way. At the end of the second day, they read a story aloud. The story included the trained nonwords; for half of the nonwords, the specific meaning could be readily inferred from context whereas for the other nonwords, only general information about the meaning of the new word could be inferred. We then assessed semantic learning using a nonword-picture matching task. Six days later (Day 3), we re-assessed semantic learning and presented orthographic choice and spelling tasks. Finally, children re-read the story for a second time, but this time silently. Following this, semantic learning was assessed for a third time.

(a) *Training.* Children were trained to pronounce 20 written nonwords such as *mouge* (see Appendix A for nonwords). Stimuli were selected from Bailey et al. (2004). They assigned each nonword a consistent pronunciation and an inconsistent pronunciation. The consistent pronunciation for each nonword corresponded to the most frequent pronunciation of the vowel grapheme. The inconsistent pronunciation corresponded to an infrequent or nonexistent pronunciation of the vowel grapheme. For each child, each nonword was assigned either a consistent (e.g., to rhyme with *gouge*) or an inconsistent (e.g., to rhyme with *rouge*) pronunciation. All nonwords were presented in both pronunciation conditions, so approximately half of the children learned a consistent pronunciation for each nonword, and the other half learned its inconsistent pronunciation. As there were 15 children in each group, consistency could not be counterbalanced exactly within groups. Rather, seven or eight children from each group learned to pronounce each nonword in each way. Training was conducted on Day 1 and Day 2 of the experiment. In the first session on Day 1 each child was familiarized with the nonwords. Nonwords were presented five at a time on laminated A4 cards in Comic Sans MS font, font size 60. The experimenter read each nonword while pointing to its orthography and asked the child to repeat it. The child then completed some filler tasks and at the end of the session, they were asked to read the list of nonwords aloud. We term this the first training trial. Errors on this and all following training trials were recorded. If a child made an error, their pronunciation was always corrected. Later on Day 1, Children completed a second session. They were asked to read the list of nonwords three times (training trials 2-4). On the following day (Day 2), each child was familiarized with the nonwords for a second time. After completing filler tasks, each child was asked to read the list of nonwords three times (training trials 5-7). Thus, all children completed two familiarization trials and seven training trials. The aim was for children to learn to pronounce all nonwords correctly by the end of the training phase. Therefore if a child could not read all nonwords correctly after the seven training trials, they were given further training (list reading with correction) until they met a criterion of being able to read the entire list of nonwords correctly. The minimum number of training trials was seven and the maximum was 13.

(b) *Story reading.* We constructed a set of four texts to make up a story about a girl called Nim. The story described events and activities throughout one day. In each text five of the trained nonwords appeared once. Thus each of the 20 nonwords appeared once in the story as a whole. Nonwords were assigned a meaning analogous to a word that children would know, for example, *mouge* was assigned to zebra (see Appendix A). Each nonword was presented in a sentence that provided cues to its meaning. Two versions of each text were constructed. In one version the five nonwords were embedded in a sentence providing a ‘helpful’ context, in the other version the sentence contexts were ‘unhelpful’ (see Appendix B for examples of helpful and unhelpful texts). The ‘helpful’ context included cues to a nonword’s exact meaning (e.g., zebra), whereas the unhelpful context provided only general cues (e.g., an animal found in Africa). Helpful and unhelpful versions of each text were developed using a cloze procedure. This ensured that nonword meanings were highly predictable in the helpful context, but minimally predictable in the unhelpful context. Helpful and unhelpful texts were matched for number of words and readability and presented individually on a laminated card in Comic Sans MS font, font size 26.

Each child was presented with a story including two helpful texts and two unhelpful texts. Stories were counterbalanced such that each nonword appeared in a helpful and unhelpful context, in each pronunciation. Thus, four different versions of the story were constructed in total, and these were counterbalanced across poor comprehenders and controls. Each child read the same story on two separate occasions: initially aloud after training on day 2 (when errors in nonword pronunciation were corrected), and then again silently six days later (Day 3).

(c) *Orthographic learning.* Children performed two tasks to assess orthographic learning on Day 3: spelling and orthographic choice. In the spelling task, children spelled the nonwords to dictation. Immediately following this they completed the orthographic choice task. The trained nonwords were presented one at a time alongside a foil on a computer screen; foil items contained the most obvious spelling pattern for the exceptional pronunciation of each nonword (for example, the target *mouge* was paired with the foil *mooge*). Children were instructed to indicate the correct spelling using a button press. Stimulus presentation was controlled by the E-Prime program (Schneider et al., 2002a, 2002b) which randomized order of presentation and recorded the speed and accuracy of children's responses.

(d) *Semantic learning.* Semantic learning was assessed at three time points: after reading the story aloud on Day 2 (SL1), following a six day delay on Day 3 (SL2) and after the second (silent) reading of the story on Day 3 (SL3). We used a nonword-picture matching task to assess semantic learning. Sets of 20 frames were constructed with each frame comprising a target nonword and four pictures. One of the pictures corresponded to the meaning implied by the helpful context (e.g., zebra); the other three pictures were appropriate alternatives for the unhelpful context (e.g., animals you might find in Africa: giraffe, elephant, lion). Following a practice trial, children were asked to indicate what they thought each word meant by pressing a button. Frames were presented in random order, using E-Prime. Accuracy scores were recorded.

Summary of primary measures from experiment. The number of trials to learn nonwords was recorded as a measure of the ability to make associations between phonology and orthography (orthographic learning). Approximately one week later, spelling and orthographic choice scores were recorded as additional measures of orthographic learning. Semantic learning was assessed at three time points; immediately after the first exposure to nonwords in context (SL1), after a six day delay (SL2) and after an additional opportunity to read the story silently (SL3). See Table 2 for a summary of the procedure for the word learning experiment.

Results

Training

The mean number of trials taken for poor comprehenders to learn inconsistent nonwords (*M* = 7.47, *SE* = .72) was higher than for consistent nonwords (*M* = 5.87, *SE* = .61). The pattern was similar for controls (inconsistent: *M* = 6.60, *SE* = .46, consistent: *M* = 5.20, *SE* = .56). An ANOVA was conducted with consistency (consistent vs. inconsistent) as a related samples factors and comprehension group (poor comprehenders vs. controls) as an independent samples factor. The ANOVA was conducted by subjects (F_s) and by items (F_i). This yielded a significant main effect of consistency ($F_s(1,28) = 13.78$, $p = .001$, $\eta^2 = .33$; $F_i(1, 38) = 11.54$, $p < .01$, $\eta^2 = .23$). Contrary to our predictions neither the main effect of group ($F_s(1,28) = 1.08$, $p = .31$, $\eta^2 = .04$; $F_i <1$) nor the interaction between group and consistency were significant (F_s and $F_i < 1$).

All children completed a minimum of seven training trials. Figure 1 shows the percentage of consistent and inconsistent items correct for each group over the first seven trials and thus provides some indication of the learning process itself. Figure 1 shows that poor comprehenders and controls learned nonwords at a similar rate but that inconsistent items took longer to learn than consistent items. An ANOVA with training trial (1 vs. 2 vs. 3 vs. 4 vs. 5 vs.6 vs. 7) and consistency (consistent vs. inconsistent) as related samples factors and group (poor comprehenders vs. controls) as an independent samples factor confirmed these observations. There was a significant main effect of training trial ($F_s(6,168) = 191.63$, $p < .001$, $\eta^2 = .87$; $F_i(6,266) = 44.25$, $p < .001$, $\eta^2 = .50$), consistency ($F_s(1,28) = 72.30$, $p < .001$, $\eta^2 = .72$; $F_i(1,266) = 78.76$, $p < .001$, $\eta^2 = .23$) and a significant interaction between trial and consistency ($F_s(6,168) = 25.34$, $p < .001$, $\eta^2 =$

.48; $F_i(6,266) = 6.73$, $p < .001$, $\eta^2 = .13$). In these analyses the main effect of comprehension group was significant by items ($F_i(1,266) = 11.50$, $p = .001$, $\eta^2 = .04$) but not by subjects $F_s(1,28) = .84$, $p > .05$, $\eta^2 = .03$). None of the interactions with comprehension group were significant (all F values < 1).

Insert Figure 1 about here

Assessing orthographic learning

Preliminary analyses suggested that there was no significant main effect of context (helpful vs. unhelpful) on spelling or orthographic choice performance (all Fs < 1 by subjects and by items). Further, context did not significantly interact with group ($F_s(1,28) = 1.74$, $p = .20$, $\eta^2 = .06$; $F_i(1,76) = 1.70$, $p = .20$, $\eta^2 = .02$) or consistency (all Fs < 1). As there was no effect of context on orthographic learning measures, for simplicity data were collapsed across context condition for the following analyses.

Spelling. Figure 2 shows that spelling was generally more accurate for consistent than inconsistent items. An ANOVA with consistency (consistent vs. inconsistent) as related samples factors and comprehension group (poor comprehenders vs. controls) as an independent samples factor confirmed this observation, yielding a significant main effect of consistency ($F_s(1,28) = 29.08$, $p < .001$, $\eta^2 = .51$; $F_i(1,38) = 14.17$, $p = .001$, $\eta^2 = .27$). In the by items analysis there was a significant main effect of comprehension group ($F_i(1,38) = 5.23$, $p = .03$, $\eta^2 = .12$), however this was not significant by subjects ($F_s(1,28) = 1.72$, $p > .20$, $\eta^2 = .06$). The interaction between group and consistency was not significant in either analysis ($F_s(1,28) = 1.37$, $p = .25$, $\eta^2 = .05$; $F_i(1,38) = 1.62$, $p = .21$, $\eta^2 = .04$).

Insert Figure 2 about here

Orthographic choice. As many children performed at or towards ceiling, our analyses focused on reaction time (RT) data. RTs (for correct responses only) were trimmed such that RTs more than two standard deviations outside of each child's mean were discarded (5.17% of RTs). Performance was generally slower for the inconsistent than the consistent items, as shown in Figure 3. An ANOVA was conducted with consistency (consistent vs. inconsistent) as a related samples factor and comprehension group (poor comprehenders vs. controls) as an independent samples factor. There was a significant main effect of consistency ($F_s(1,28) = 8.23$, $p < .01$, $\eta^2 = .23$; $F_i(1,38) = 5.89$, $p = .02$, $\eta^2 = .13$). There was a trend for a main effect of comprehension group in the analysis by items ($F_i(1,38) = 3.51$, $p = .07$, $\eta^2 = .09$), however this was not the case by subjects ($F_s < 1$). The interaction between group and consistency was not significant in either analysis ($F_s(1,28) = 1.16$, $p = .29$, $\eta^2 = .04$; $F_i < 1$).

Insert Figure 3 about here

Assessing semantic learning

For the semantic learning task children were required to choose each target from an array of four pictures. As this dependent variable is bounded, raw scores were subjected to an angular transformation (as recommended by Kirk, 1968, p. 66). Note though that analyses conducted with raw scores yielded an identical pattern of results. Preliminary analyses suggested that there was no significant effect of consistency on semantic learning and consistency did not significantly interact with any other variable. Notably, there was no main effect of consistency nor interaction with context or comprehension group (*F*s < 1). Therefore, data were collapsed across consistency. Figure 4 shows mean performance of poor comprehenders and controls on the semantic learning tasks in helpful and unhelpful conditions. Panel (a) shows retention over a six-day delay (SL1 vs. SL2) and panel (b) shows change in performance with re-exposure to nonwords in story context. The dashed lines in both panels of Figure 4 correspond to performance at chance level (0.25). It is clear that children performed better in the helpful condition and that controls showed greater accuracy in all conditions. Over the six-day retention interval (SL1 vs. SL2; top panel of Figure 4), the control group remained relatively constant whereas poor comprehenders' performance declined in the helpful condition. For both groups, re-exposure to nonwords in story context (SL1 vs. SL3; bottom panel of Figure 4) enhanced performance in the helpful condition.

Insert Figure 4 about here

Two sets of analysis were conducted to investigate retention and exposure effects on semantic learning. The first ANOVA included testing time (SL1 vs. SL2) and context (helpful vs. unhelpful) as related samples factors and comprehension group (poor comprehenders vs. controls) as an independent samples factor. Consistent with observations above, there were significant main effects of context ($F_s(1,28) = 23.08$, $p < .001$, $\eta^2 = .45$; $F_i(1, 76) = 22.18$, $p < .001$, $\eta^2 = .23$) and comprehension group ($F_s(1,28) = 11.53$, $p < .01$, $\eta^2 = .29$; $F_i(1, 76) = 18.55$, $p < .001$, $\eta^2 = .20$). The analysis also revealed a significant interaction between testing time and context in the by subjects analysis ($F_s(1,28) = 6.58$, $p < .05$, $\eta^2 = .19$). Figure 4 suggests that this reflects a greater effect of context in SL1, immediately after exposure to nonwords. Note however that this effect was not significant by items ($F_i(1, 76) = 2.11$, $p > .05$, $\eta^2 = .03$).

The second ANOVA included testing time (SL1 vs. SL3) and context (helpful vs. unhelpful) as related samples factors and comprehension group (poor comprehenders vs. controls) as an independent samples factor. Again there were significant main effects of context ($F_s(1,28) = 52.70$, $p < .001$, $\eta^2 = .65$; $F_i(1, 76) = 55.12$, $p < .001$, $\eta^2 = .42$) and comprehension group ($F_s(1,28) = 9.16$, $p < .01$, $\eta^2 = .25$; $F_i(1, 76) = 16.24$, $p < .001$, $\eta^2 = .18$). In this analysis there was also a significant main effect of testing time ($F_s(1,28) = 24.07$, $p < .001$, $\eta^2 = .46$; $F_i(1, 76) = 7.29$, $p < .01$, $\eta^2 = .09$). Figure 4 suggests that this reflects improved performance in SL3, after a second exposure to nonwords. The analysis also revealed a significant interaction between testing time and context ($F_s(1,28) = 11.08$, $p < .01$, $\eta^2 = .28$; $F_i(1, 76) = 4.43$, $p < .05$, $\eta^2 = .06$). In this case the interaction was significant in both subjects and items analyses. Figure 4 suggests that this interaction reflects greatly improved performance in the helpful condition of SL3.

It is important to note that in the unhelpful condition performance should be at chance levels as the stories did not provide adequate information for correct performance. However, controls performed above chance at all testing points whereas poor comprehenders tended to perform below chance, significantly so in SL1 ($t(14) = -2.62$, $p < .05$)[2]. It is not clear why this is the case and exploratory analyses of responses on this task did not yield any systematic biases, but this does explain why group differences are manifested as a main effect rather than the expected interaction.

Post-hoc *t*-tests were conducted to investigate critical group differences in the helpful condition of SL1, SL2 and SL3. This yielded a significant difference between groups in SL2 ($t(28) = -2.44$, $p = 0.02$), but not in SL1 ($t(28) = -1.27$, $p = 0.22$) or SL3 ($t(28) = -1.76$, $p = 0.09$). This suggests that poor comprehenders and controls benefited from context to the same extent in SL1 - immediately after the first exposure to nonwords. The group difference in SL2 suggests that controls were better than poor comprehenders at retaining semantic information in the six days between SL1 and SL2. As described above, there was a main effect of testing time between SL1 and SL3, suggesting that all children benefited from the additional exposure to nonwords in context. There was a trend for a group difference in the helpful condition of SL3, suggesting that perhaps poor comprehenders benefited less; however this finding was not significant.

Discussion

The aim of this experiment was to investigate whether poor comprehenders have difficulty making associations between a novel word's phonology and its orthographic and/or semantic form – we term these processes orthographic learning and semantic learning respectively. We also investigated orthographic and semantic learning for items with consistent and inconsistent spellings. We will discuss the results for orthographic and semantic learning measures in turn.

We measured orthographic learning in terms of the number of trials taken to learn nonwords, performance across the training trials and using two post-tests: spelling and orthographic choice. Across all measures, performance was higher for consistent versus inconsistent nonwords suggesting – unsurprisingly – that inconsistency in spelling-sound mappings poses a problem for word learners. However, we found no clear evidence that poor comprehenders had difficulty learning consistent or inconsistent nonwords. In addition, we did not observe the predicted interaction between group and consistency.

Across all measures poor comprehenders did not differ from controls in the by-subjects analyses. However, as there was some evidence for a group trend in the items analyses, it seems prudent not to over interpret this finding and it clearly needs replicating with a larger sample.

In light of this trend, it is worth noting that consistency for some items was questionable[2]. Following Bailey et al. (2004), pronunciations were classified as consistent if their vowel grapheme was pronounced in the most frequent way. However, consistency is a complex issue and there are different ways of defining it. Many argue that consistency is continuous rather than binary. One reason for this is that the frequency of grapheme pronunciations varies in a graded manner according to both type and token frequency. Another important issue is that a vowel's consistency is often conditional on the phonemes that surround it. Kessler and Treiman (2001) found that vowel pronunciation was particularly determined by subsequent phonemes, suggesting that defining consistency at the level of the rime might be helpful. However, other surrounding phonemes also appear to play a role (for a fuller discussion of these issues see Caravolas, Kessler, Hulme & Snowling, 2005; Kessler & Treiman, 2001). Given this complexity, a more sophisticated treatment of consistency when selecting items is warranted. Potentially, this could reveal deficits in poor comprehenders' learning of inconsistent/exceptional novel words.

We turn now to discuss semantic learning. Many studies have shown that poor comprehenders have difficulty with semantic tasks (e.g., Cain et al., 2003; Cain et al., 2004; Nation & Snowling, 1998). Consistent with these studies, our poor comprehenders had poorer existing oral vocabulary knowledge than controls. In terms of novel word learning, there was a main effect of context such that all children used the cues in the helpful condition to infer the meaning of nonwords. Also, across all analyses there was a main effect of comprehension group, with controls outperforming poor comprehenders. This result is difficult to interpret given the unexpected group difference in the unhelpful condition. This finding warrants further investigation. Nevertheless, our results show that poor comprehenders were less-skilled at inferring semantic information from the helpful context, and that they were significantly worse at retaining semantic information over time.

On the basis of previous work, we had predicted that poor comprehenders would show weaker semantic learning than controls at all time points. Interestingly however, the most robust evidence for reduced semantic learning was seen after a delay. One reason for this may be the nature of our task. We assessed semantic learning using a nonword-picture matching task whereas Cain et al. (2003; 2004) asked children to provide definitions of learned nonwords. Plausibly, poor comprehenders' semantic learning abilities are sufficient to support the recognition processes necessary to immediately perform a nonword-picture matching task (as in this study), but not the more demanding task of producing of a definition. Consistent with this, Nation et al. (2007) found that poor comprehenders produced fewer correct definitions to recently learned new words than control children, but performed well on a simpler recognition task. Interestingly, Nation et al. also observed a fall-off in learning over time, with group differences emerging on the simpler tasks one week later, mirroring the effects we observed in our experiment. Retention weaknesses in poor comprehenders have also been reported elsewhere (Cragg & Nation, 2006; for more information on memory weaknesses in poor comprehenders see Cain, 2006; Nation, Adams, Bowyer-Crane & Snowling, 1999; Weekes, Hamilton, Oakhill & Holliday, 2007). Taken together, these studies are consistent with the notion that poor comprehenders can make some links between words and their semantic properties, but that they have difficulty developing rich and durable semantic representations.

It is also worth noting another major difference between this and previous studies exploring new vocabulary learning in poor comprehenders (Cain et al., 2003; Cain et al., 2004; Nation et al., 2007). Our children learned to associate new phonological and orthographic forms with known objects rather than novel objects or concepts. Therefore they were not learning new semantic information but rather novel names for easily nameable objects, akin to second language learning. Potentially, this may have underestimated true semantic learning. We had intended that children would form associations between the new word and the semantic information provided by the text. However, the children may have done something different: they might have formed associations between the novel phonological/orthographic form and the known phonological/orthographic form (i.e., after realizing that a *mouge* was a zebra, they then formed an association between the word *mouge* and the word zebra). Arguably, this strategy might have been sufficient to perform the nonword-picture matching task. It would be interesting to examine how well poor comprehenders would perform our task, if this scaffold from existing knowledge was removed. We anticipate that group differences between skilled and less-skilled comprehenders would widen, if new referents or concepts for new forms had to be learned.

Before closing, it is important to note that although poor comprehenders did not have difficulty learning the orthography of inconsistent items in this experiment, they did show an exception word reading deficit, consistent with other studies (Nation & Snowling, 1998; Ricketts et al., 2007). The question arises then of what might cause poor comprehenders' exception word reading difficulties. While our study does not provide an opportunity to test causal hypotheses, our data are consistent with the view that poor comprehenders' difficulty with reading exception words is a manifestation of their underlying vocabulary weaknesses, as proposed by Nation and Snowling (1998). Put simply, if a child does not have an exception word in their oral repertoire, they will not be well-placed to read it correctly as they will have no basis from which to bring their partial decoding attempt "in-line" with the correct pronunciation. When provided with a pronunciation (as in this experiment), they are perfectly adept at learning it, even if associations between orthography and phonology are inconsistent. However, without this external teacher, they are less able to bring their own semantic knowledge to the task of assigning a pronunciation to a novel string. While this does not have an effect on their ability to read regular words, it is detrimental to exception word reading.

In conclusion, as reading provides an opportunity to learn new words, children with poor reading comprehension may not be able to capitalize on this strategy as much as their more skilled peers, impacting any initial vocabulary weaknesses. At the same time, and as discussed by the Lexical Quality Hypothesis (Perfetti & Hart, 2002), developing good quality word-level representations is the fundamental foundation to reading comprehension. Thus, the relationship between reading comprehension and word learning is likely to be reciprocal, with each set of skills scaffolding the other through development. Our findings highlight the need to consider both semantic and orthographic aspects of lexical development if we are to fully account for how children develop representations of the quality needed to underpin skilled reading.

References

Bailey, C. E., Manis, F. R., Pedersen, W. C., & Seidenberg, M. S. (2004). Variation among developmental dyslexics: Evidence from a printed-word-learning task. *Journal of Experimental Child Psychology, 87*, 125-154.

Bowey, J. A., & Rutherford, J. (2007). Imbalanced word-reading profiles in eighth-graders. *Journal of Experimental Child Psychology, 96*(3), 169-196.

Bowyer-Crane, C. A., & Snowling, M. J. (2005). Assessing children's inference generation: What do tests of reading comprehension measure? *British Journal of Educational Psychology, 75*(2), 189-201.

Burt, J. S. (2006). What is orthographic processing skill and how does it relate to word identification in reading? *Journal of Research in Reading, 29*(4), 400-417.

Byrne, B., Freebody, P., & Gates, A. (1992). Longitudinal data on the relations of word-reading strategies to comprehension, reading time, and phonemic awareness. *Reading Research Quarterly, 27*(2), 140-151.

Cain, K. (2003). Text comprehension and its relation to coherence and cohesion in children's fictional narratives. *British Journal of Developmental Psychology, 21*, 335-351.

Cain, K. (2006). Individual differences in children's memory and reading comprehension: An investigation of semantic and inhibitory deficits. *Memory, 14*(5), 553-569.

Cain, K., & Oakhill, J. (2007). Reading comprehension difficulties: Correlates, causes and consequences. In K. Cain & J. Oakhill (Eds.), *Children's Comprehension*

Problems in Oral and Written Language: A Cognitive Perspective. (pp. 41-75): Guildford Press.

Cain, K., Oakhill, J., & Elbro, C. (2003). The ability to learn new word meanings from context by school-age children with and without language comprehension difficulties. *Journal of Child Language, 30*, 681-694.

Cain, K., Oakhill, J., & Lemmon, K. (2004). Individual differences in the inference of word meanings from context: The influence of reading comprehension, vocabulary knowledge, and memory capacity. *Journal of Educational Psychology, 96*, 671-681.

Caravolas, M., Kessler, B., Hulme, C., & Snowling, M. (2005). Effects of orthographic consistency, frequency, and letter knowledge on children's vowel spelling development. *Journal of Experimental Child Psychology, 92*(4), 307-321.

Castles, A., & Holmes, V. (1996). Subtypes of developmental dyslexia and lexical acquisition. *Australian Journal of Psychology, 48*(3), 130-135.

Castles, A., & Nation, K. (2006). How does orthographic learning happen? In S. Andrews (Ed.), *From inkmarks to ideas: Challenges and controversies about word recognition and reading*. Hove, East Sussex: Psychology Press.

Catts, H. W., Adlof, S. M., & Weismer, S. E. (2006). Language deficits in poor comprehenders: A case for the simple view of reading. *Journal of Speech, Language, and Hearing Research, 49*, 278-293.

Coltheart, M., & Leahy, J. (1996). Assessment of lexical and nonlexical reading abilities in children: Some normative data. *Australian Journal of Psychology, 48*, 136-140.

Cragg, L., & Nation, K. (2006). Exploring written narrative in children with poor reading comprehension. *Educational Psychology, 26*(1), 55-72.

Harm, M., & Seidenberg, M. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review, 106*(3), 491-528.

Harm, M., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review, 111*(3), 662-720.

Keenan, J. M., & Betjemann, R. S. (2007). Comprehension of single words: The role of semantics in word identification and reading disability. In E. Grigorenko & A. Naples (Eds.), *Single-word reading: Behavioral and biological perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates.

Kessler, B., & Treiman, R. (2001). Relationships between sounds and letters in English monosyllables. *Journal of Memory and Language, 44*, 592-617.

Kirk, R.E. (1968). Experimental design: procedures for the behavioural sciences. Belmont, CA: Brooks/Cole.

Manis, F. R., Seidenberg, M., Doi, L. M., McBride-Chang, C., & Petersen, A. (1996). On the bases of two subtypes of developmental dyslexia. *Cognition, 58*, 157-195.

Masterson, J., Dixon, M., & Stuart, M. (2002). *The Children's Printed Word Database,* Retrieved 8 November 2005 from http://www.essex.ac.uk/psychology/cpwd/

McDougall, S., Hulme, C., Ellis, A., & Monk, A. (1994). Learning to read: The role of short-term memory and phonological skills. *Journal of Experimental Child Psychology, 58*, 112-133.

Nation, K. (2005). Children's reading comprehension difficulties. In M. J. Snowling, C. Hulme & M. Seidenberg (Eds.), *The science of reading: A handbook* (pp. 248-265). Oxford: Blackwell.

Nation, K., Adams, J. W., Bowyer-Crane, C. A., & Snowling, M. J. (1999). Working memory deficits in poor comprehenders reflect underlying language impairments. *Journal of Experimental Child Psychology, 73*, 139-158.

Nation, K., Clarke, P., Marshall, C. M., & Durand, M. (2004). Hidden language impairments in children: Parallels between poor reading comprehension and specific language impairment. *Journal of Speech, Language, and Hearing Research, 47*, 199-211.

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*, 85-101.

Nation, K., & Snowling, M. J. (2004). Beyond phonological skills: Broader language skills contribute to the development of reading. *Journal of Research in Reading, 27*, 342-356.

Nation, K., Snowling, M. J., & Clarke, P. (2007). Dissecting the relationship between language skills and learning to read: Semantic and phonological contributions to new vocabulary learning in children with poor reading comprehension. *Advances in Speech-Language Pathology, 9*, 131-139.

Neale, M. (1997). Neale Analysis of Reading Ability - Second Revised British Edition. Windsor: NFER.

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro & P. Reitsma (Eds.), *Precursors of functional literacy* (Vol. 11). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Plaut, D. C., McClelland, J. L., Seidenberg, M., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103*, 56-115.

Ricketts, J., Nation, K., & Bishop, D. V. (2007). Vocabulary is important for some, but not all reading skills. *Scientific Studies of Reading*, *11*, 235-257

Schneider, W., Eschman, A., & Zuccolotto, A. (2002a). *E-Prime reference guide*. Pittsburgh: Psychology Software Tools Inc.

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

Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition, 55*, 151-218.

Stothard, S. E., & Hulme, C. (1992). Reading comprehension difficulties in children: The role of language comprehension and working memory skills. *Reading and Writing, 4*, 245-256.

Torgesen, J., Wagner, R., & Rashotte, C. (1999). *Test of Word Reading Efficiency*. Austin, TX: Pro-Ed.

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence*. London: The Psychological Corporation.

Weekes, B. S., Hamilton, S., Oakhill, J. V., & Holliday, R. E. (2007). False recollection in children with reading comprehension difficulties. *Cognition*.

Wiig, E. H., & Secord, W. (1992). *Test of Word Knowledge*. US: The Psychological Corporation.

Author notes

Jessie Ricketts, D.V.M. Bishop, Kate Nation, University of Oxford. We would like to thank The Economic and Social Research Council for a studentship awarded to the first author and grant awarded to the second and third authors. The third author is supported by The Wellcome Trust. We would also like to thank all teachers, parents and children for their ongoing support with our research.

Correspondence should be addressed to Jessie Ricketts, Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford, OX1 3UD, UK. Email: jessie.ricketts@psy.ox.ac.uk; Web: www.psy.ox.ac.uk/lcd, www.psy.ox.ac.uk/oscci

Footnotes

[1] Vocabulary knowledge (WASI; Wechsler, 1999) and exception word reading (word list from Coltheart and Leahy, 1996) were also assessed during the screening phase 10 months prior to the experiment. At this time point poor comprehenders achieved significantly lower scores on both measures (see Ricketts, Nation & Bishop, 2007) replicating the data presented here.

[2] We thank an anonymous reviewer for noting that for some of our items (duite, froupe, grast, mouge, smune, tauge), the relationship between their spelling and the 'consistent' pronunciation assigned to them by Bailey et al. (2004) is more ambiguous. Data from the Children's Printed Word Database (CPWD; Masterson, Dixon & Stuart, 2002) supports this observation. The reviewer suggested that by removing these potentially problematic stimuli we might observe a robust group effect or a group by consistency interaction on orthographic learning. However, analyses conducted with a reduced set of 14 stimuli yielded a pattern of results for the training, spelling and orthographic choice measures that was identical to that described above. Note that data from the CPWD show that for these 14 items, the 'consistent' pronunciation would be the most common pronunciation in British English.

Table 1.

Mean performance of poor and skilled comprehension groups on selection and background measures

	Poor Comprehenders (PC, N=15)		Controls (N=15)		PC vs. controls *F*(1,28)	Estimated effect size
	M	*SD*	*M*	*SD*		
<u>Screening measures</u>						
Chronological age[1]	9.21	0.30	9.26	0.28	0.22	0.01
WASI Matrices[2]	52.33	5.02	52.33	4.29	0.00	0.00
TOWRE decoding[3]	107.67	13.11	108.27	9.68	0.02	0.00
NARA-II Reading comprehension[3]	81.93	2.69	103.13	4.88	217.14***	0.89
NARA-II Text reading accuracy[3]	99.27	9.43	108.07	8.18	7.45*	0.21
<u>Background measures</u>						
Chronological age[1]	9.99	0.28	10.07	0.34	0.54	0.02
Phoneme deletion[4]	0.75	0.11	0.77	0.14	0.11	0.00
TOWK vocabulary[5]	8.00	2.14	10.87	1.51	18.03***	0.39
Exception word reading[4]	0.75	0.13	0.87	0.06	10.11**	0.27

Notes. * p<0.05; ** p< 0.01, *** p< 0.001; [1]In years; [2]T-score, M = 50, SD = 10; [3]Standard scores, M = 100, SD = 15; [4]Proportion correct; [5]Scaled scores, M = 10, SD = 3

Table 2.

Summary of procedure for word learning experiment

	Day 1	Day 2	Day 3
AM	Nonword familiarization	Nonword familiarization	Orthographic learning
	Filler tasks	Filler tasks	1. Spelling
	Training trial 1	Training trials 5, 6 and 7	2. Orthographic choice
		Further training if necessary	SL2
			Filler tasks
			Story reading 2 (silent)
			SL3
PM	Training trials 2, 3 and 4	Story reading 1 (aloud)	
		SL1	

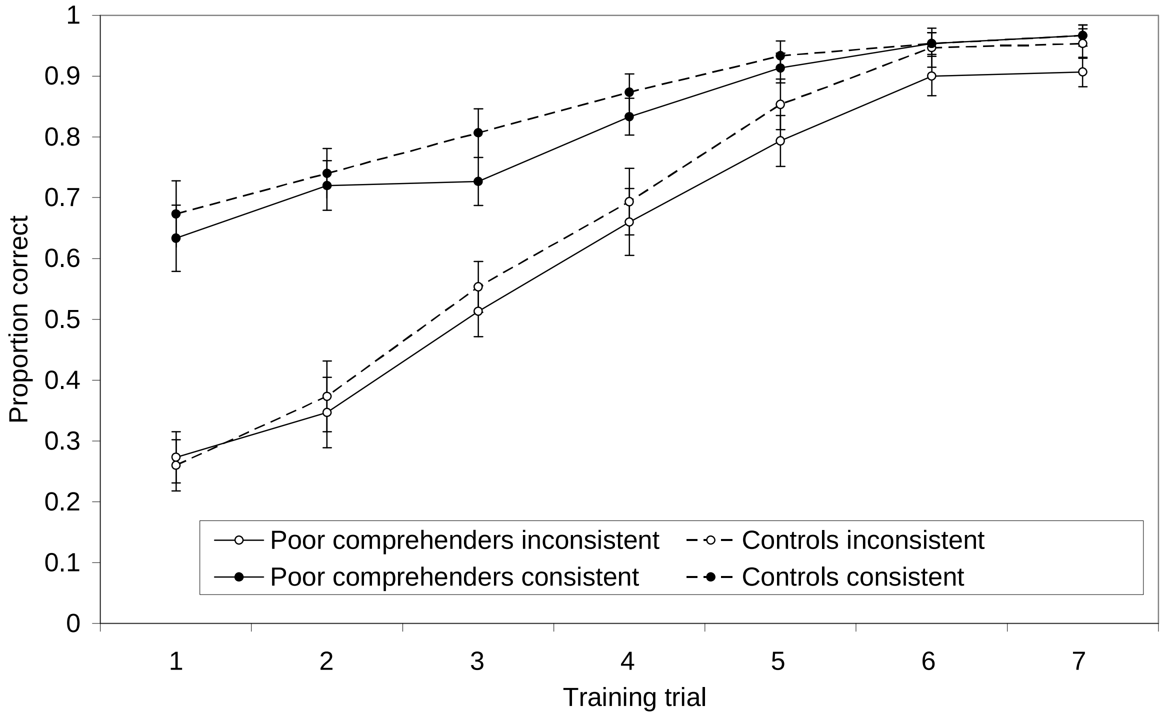
Figure Captions

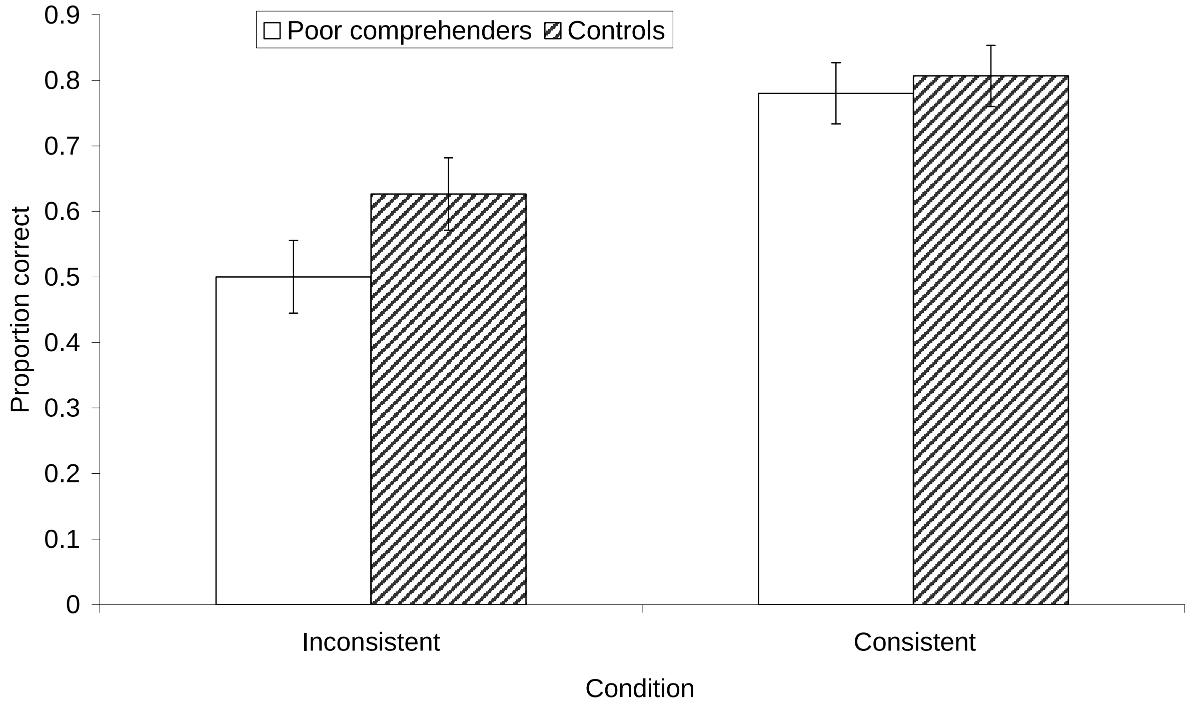
Figure 1. Mean proportion (±*SE*) of consistent and inconsistent items correct for each group across the first seven training trials.

Figure 2. Mean proportion correct in the spelling task (±*SE*) for poor comprehender and control groups for consistent and inconsistent items.

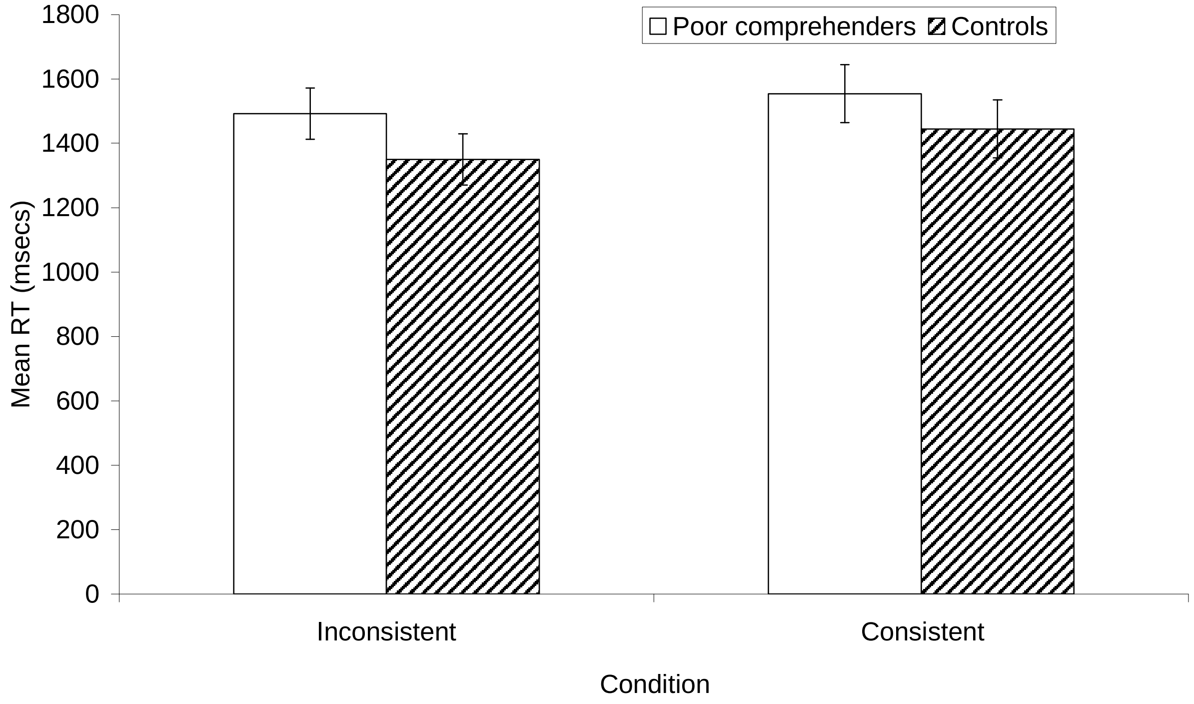
Figure 3. Grand mean reaction time data in the orthographic choice task (±*SE*) for poor comprehender and control groups for consistent and inconsistent items.

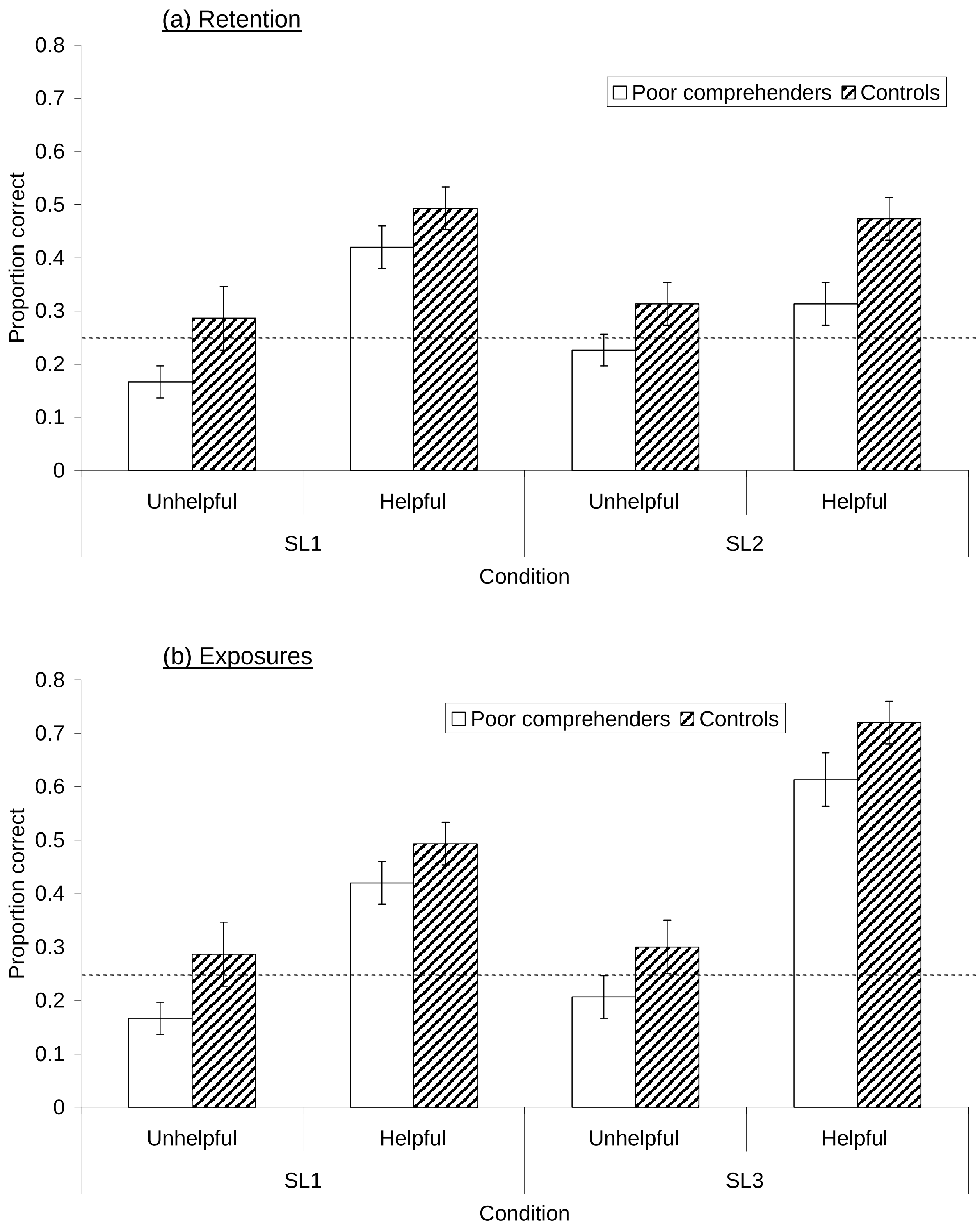
Figure 4. Mean proportion correct (±SE) in semantic learning tasks (SL1, SL2 and SL3) for poor comprehender and control groups in unhelpful and helpful conditions. Panel (a) demonstrates retention over time (SL1 vs. SL2); panel (b) demonstrates change in performance with re-exposure to nonwords in story context (SL1 vs. SL3). The dashed lines show chance level performance.


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Proportion correct
1
2
3
4
5
6
7
Training trial
Poor comprehenders inconsistent
Controls inconsistent
Poor comprehenders consistent
Controls consistent



0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Proportion correct
Poor comprehenders
Controls
Inconsistent
Consistent
Condition


1800
1600
1400
1200
1000
800
600
400
200
0
Mean RT (msecs)
Poor comprehenders
Controls
Inconsistent
Consistent
Condition


(a) Retention
Poor comprehenders
Controls
Proportion correct
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Unhelpful
Helpful
Unhelpful
Helpful
SL1
SL2
Condition
(b) Exposures
Poor comprehenders
Controls
Proportion correct
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Unhelpful
Helpful
Unhelpful
Helpful
SL1
SL3
Condition