Listening while reading promotes word learning from stories

Journal of Experimental Child Psychology 167 (2018) 10–31

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

Listening while reading promotes word learning from stories

Alessandra Valentini [a,*], Jessie Ricketts [b], Rachel E. Pye [c], Carmel Houston-Price [a]

[a] School of Psychology and Clinical Language Sciences, University of Reading, Reading RG6 6AL, UK
[b] Department of Psychology, Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK
[c] School of Psychology and Clinical Language Sciences, University of Reading Malaysia, 79200 Iskandar Puteri, Johor, Malaysia

ARTICLE INFO

Article history:
Received 4 November 2016
Revised 13 September 2017
Available online 15 November 2017

Keywords:
Vocabulary acquisition
Reading
Reading while listening
Listening to stories
Phonology
Orthography

ABSTRACT

Reading and listening to stories fosters vocabulary development. Studies of single word learning suggest that new words are more likely to be learned when both their oral and written forms are provided, compared with when only one form is given. This study explored children's learning of phonological, orthographic, and semantic information about words encountered in a story context. A total of 71 children (8- and 9-year-olds) were exposed to a story containing novel words in one of three conditions: (a) listening, (b) reading, or (c) simultaneous listening and reading ("combined" condition). Half of the novel words were presented with a definition, and half were presented without a definition. Both phonological and orthographic learning were assessed through recognition tasks. Semantic learning was measured using three tasks assessing recognition of each word's category, subcategory, and definition. Phonological learning was observed in all conditions, showing that phonological recoding supported the acquisition of phonological forms when children were not exposed to phonology (the reading condition). In contrast, children showed orthographic learning of the novel words only when they were exposed to orthographic forms, indicating that exposure to phonological forms alone did not prompt the establishment of orthographic representations. Semantic learning was greater in the combined condition than in the listening and reading conditions. The presence of the definition was associated with better performance on the semantic subcategory and definition posttests but not on the phonological,

* Corresponding author.
E-mail address: a.valentini@reading.ac.uk (A. Valentini).

https://doi.org/10.1016/j.jecp.2017.09.022
0022-0965/

orthographic, or category posttests. Findings are discussed in relation to the lexical quality hypothesis and the availability of attentional resources.

Introduction

Vocabulary development starts during infancy and is a lifelong endeavor; children and adults acquire new words, and specify existing lexical representations, throughout the lifespan (Nagy, Anderson, & Herman, 1987). The majority of words are not acquired through direct instruction but rather incidentally from conversations, television, and texts (Akhtar, 2004; Elley, 1989; Henderson, Devine, Weighall, & Gaskell, 2015; Houston-Price, Howe, & Lintern, 2014). The current study explored how children learn new words when they are exposed to them incidentally in stories. To our knowledge, this is the first investigation of whether children show greater word learning from listening to, reading, or both listening to and reading stories.

Many studies have shown that exposure to stories fosters vocabulary development in children (Henderson et al., 2015; Nagy et al., 1987; Ricketts, Bishop, Pimperton, & Nation, 2011; Wilkinson & Houston-Price, 2013; Williams & Horst, 2014). Suggate, Lenhard, Neudecker, and Schneider (2013) compared the word learning shown in three story presentation conditions: independent reading, listening to an adult reading the story, and listening to an adult telling the story in his own words. In the listening conditions children (8- to 10-year-olds) were exposed to the spoken forms (phonology) and meanings (semantics) of new words but not their written forms (orthography), whereas in the reading condition they encountered the words' written forms (orthography) and meanings. The children who listened to the stories were more likely to demonstrate knowledge of the new words' meanings than the children who read the stories, suggesting that oral presentation is more beneficial for vocabulary learning in school-aged children than written presentation. In contrast to this result, studies of adults learning English as a second language tend to show that participants acquire new words more easily if presented with material in written form rather than oral form (Brown, Waring, & Donkaewbua, 2008; Sydorenko, 2010). Similarly, studies exploring memory for word lists or verse show better performance for written material than for oral material in both adults and children (Hartman, 1961; Menne & Menne, 1972). Because the primary medium for vocabulary acquisition is oral language, it makes sense that oral presentation may be the preferred and easiest method for acquiring vocabulary early on, but as reading ability improves, children become better at learning from written texts.

There is reason to suppose that both listening to and reading a story at the same time will be maximally beneficial for word learning. Studies using e-book presentations show that presentation in both modalities is more beneficial for vocabulary acquisition than simply listening to the book read by an adult (Shamir, Korat, & Fellah, 2012). Other work has also found that access to orthographic forms promotes oral vocabulary learning, an effect referred to as "orthographic facilitation" (e.g., Hu, 2008; Ricketts, Bishop, & Nation, 2009; Ricketts, Dockrell, Patel, Charman, & Lindsay, 2015; Rosenthal & Ehri, 2008). In these studies, children were taught phonological forms and semantics either with or without orthography; greater learning of phonology, semantics, and orthography was seen for items where orthography was provided. To date, studies investigating orthographic facilitation all have employed a direct instruction approach to teaching new words. Whether these findings generalize to an incidental learning context, where children's attention is not explicitly drawn to the new words, needs to be explored.

Further evidence that simultaneously listening to and reading stories leads to better learning than a single modality of presentation comes from a study by Rosenthal and Ehri (2011). In that study, children read stories that contained novel words silently, pronouncing half of the new words aloud when they encountered them. Semantic and orthographic learning was greater for words that had been pronounced, demonstrating "phonological facilitation." Although in this study words were embedded in

passages, thereby simulating incidental learning from reading, attention was drawn to the target words (through underlining), and only these words were pronounced aloud. In the current study, to provide a more realistic learning context, new words were not highlighted and presentation modality was manipulated for the story as a whole rather than for individual words.

To date, listening and reading conditions have not been compared with a combined listening and reading condition in relation to school-aged children's learning of vocabulary in their first language. However, studies of adult second-language learners typically find superior learning when material is presented in written or dual modality format rather than in spoken format (Brown et al., 2008; Sydorenko, 2010). For example, Brown et al. (2008) presented three stories in three different modalities (listening, reading, and combined) to 35 Japanese adult students and found greater semantic learning in the reading and combined conditions compared with the listening condition. In line with these results, studies investigating the impact of multimodal presentation on memory for lists of words or verse found superior performance in the combined and reading conditions compared with the listening condition in children aged 8 or 9 years and adults (Hartman, 1961; Menne & Menne, 1972).

In summary, evidence of orthographic and phonological facilitation for vocabulary acquisition has so far come from studies that present words in isolation, rather than in context (Ricketts et al., 2009; Rosenthal & Ehri, 2008), or from studies that confine the dual modality of presentation to the words of interest, rather than to the narrative as a whole (Rosenthal & Ehri, 2011). Evidence for the beneficial effect of dual presentation modality compared with single modality presentation has also come from studies with adult second-language learners (Brown et al., 2008). Therefore, questions remain about whether such facilitation effects are seen when school-aged children are learning new words in their first language and when the presentation modality is extended to the entire story in which words are embedded.

Why might there be an advantage for a combined presentation modality over a single presentation modality? One possibility is that providing both oral and written forms frees up attentional resources during encoding, meaning that resources can be allocated to story comprehension and to the encoding of new word meanings. This idea is related to Ehri's (2014) views on how good decoders use their strong knowledge of the link between orthography and phonology to generate phonology automatically while reading, leaving more resources for text comprehension and the learning of words' meanings. It also resonates with cognitive load theory in multimedia learning (Mayer, Moreno, Boire, & Vagge, 1999), which states that situations that reduce cognitive load are more conducive to learning. According to cognitive load theory, presenting information in two modalities frees resources by increasing working memory capacity. This process occurs online as new information is encountered and while a representation is being created (Mousavi, Low, & Sweller, 1995). On this view, then, presenting both phonological and orthographic forms may free cognitive resources so that more attention can be paid to comprehension and to encoding word meanings. Notably, this framework focuses on the conditions that facilitate learning, suggesting that presenting information in a dual modality creates connections between two representations. It does not, however, probe the nature of the representations created (Mayer et al., 1999).

An alternative framework for interpreting the benefits of simultaneous bimodal presentation is provided by the lexical quality hypothesis (Perfetti & Hart, 2002). This hypothesis posits that lexical representations that include both phonological and orthographic information are of "higher quality" than representations that contain only phonological or orthographic information. Although the lexical quality hypothesis focuses on existing lexical representations rather than their acquisition, it is consistent with the idea that a combined presentation will support the building of better quality lexical representations that, therefore, can be more readily accessed when they are encountered at a later stage.

In addition to story presentation modality, the content of the story and the kinds of information provided within the story are likely to play a role in children's learning of new words. While reading, children use contextual cues to infer semantic information about new words even when these cues convey minimal information (Nagy et al., 1987). Whereas general context supports learning of the broad category of a new word, more constraining contexts enhance learning of a word's specific features (Ricketts et al., 2011). In addition, when words are encountered aurally, presenting a definition

alongside the new word is beneficial for semantic learning (Dickinson, 1984; Penno, Wilkinson, & Moore, 2002; Wilkinson & Houston-Price, 2013). There are important differences in the information conveyed by contextual cues and definitions. Contextual cues typically provide general information about a word's meaning and are spread through the text (Nagy, 1995). Definitions provide detailed semantic information about the word and typically occur alongside its form. If definitions are considered a highly constraining context, they should allow more specific information about a word to be extracted than the more general categorical information revealed by less constrained contexts (Ricketts et al., 2011). Although previous studies have shown that definitions elicit greater semantic learning than general contextual support in both recognition tasks (Wilkinson & Houston-Price, 2013) and production tasks (Justice, Meier, & Walpole, 2005), previous research has not explored the nature of the semantic learning shown in such conditions, a further aim of the current study.

The current study

Children aged 8 or 9 years were exposed to a story containing eight new words. This age range was chosen to ensure a range of reading abilities in a population of children used to reading and understanding written texts. Children were divided into three groups, with one group listening to the story (listening group), one group reading the story (reading group), and one group reading and listening to the story simultaneously (combined group). Half of the words were accompanied by a definition the first time they were presented, and the other half of the words were not. The story was presented twice over 2 weeks to promote learning via repetition and allow for sleep-related consolidation (e.g., Henderson, Weighall, & Gaskell, 2013) and to fit in with the school timetable. After the second story presentation, children's knowledge of phonological forms, orthographic forms, and meanings was assessed in a series of posttests. In phonological and orthographic posttests, children were required to recognize correct spoken or written forms from two alternatives. In three semantic posttests, children were asked to identify the words' categories, subcategories, and definitions. By using these tasks, we were able to capture acquisition of different aspects of semantic information about the given words.

From a practical perspective, this investigation serves to establish how best to promote vocabulary acquisition in school-aged children and whether children with different ability levels are facilitated by different presentation modalities. From a theoretical perspective, the results assess the hypothesis that different presentation modalities facilitate the acquisition of words of higher or lower lexical quality. The framework provided by the lexical quality hypothesis (Perfetti & Hart, 2002) has previously been applied to evidence that better readers create higher quality representations of new words (Perfetti, 2007), but it has not been used to make clear predictions about the conditions under which words are optimally encoded or retained. This study explored whether word representations created through different presentation modalities vary in their lexical quality.

The study addressed several hypotheses regarding orthographic, phonological, and semantic learning. In relation to dual versus single modality presentations, it was hypothesized that the combined condition would elicit superior orthographic, phonological and semantic learning (Ricketts et al., 2009; Rosenthal & Ehri, 2008; Rosenthal & Ehri, 2011) compared with the other two conditions. Regarding semantic learning, given the conflicting evidence, clear predictions could not be made as to whether the listening group would outperform the reading group in learning words' meanings, as suggested by the first language literature (Suggate et al., 2013), or whether the reading group would show an advantage, as suggested by the second language and memory literatures (Brown et al., 2008; Menne & Menne, 1972). In line with previous research on word learning from stories, it was expected that definitions would promote semantic learning (Penno et al., 2002; Wilkinson & Houston-Price, 2013), particularly in tasks assessing learning of specific features of the words' meanings.

Individual differences might also constrain children's learning of new words from spoken and written story exposure. Therefore, measures of children's abilities were collected. Based on previous research, it was expected that reading accuracy would predict learning of orthographic forms in the reading condition (Ricketts et al., 2011) and that vocabulary knowledge, reading accuracy, and reading comprehension would predict semantic learning in the reading condition (Cain, Oakhill, & Lemmon, 2004; Ricketts et al., 2011), whereas oral vocabulary ability should predict semantic learning in the

listening condition (Lin, 2014; Penno et al., 2002; Senechal, Thomas, & Monker, 1995). Given that previous studies have not investigated monolingual children's word learning when children are simultaneously reading and listening, it was unclear which abilities might influence learning in the combined condition. We anticipated that because children could rely on both oral and written modalities, oral vocabulary, reading accuracy, and reading comprehension all would influence performance in the combined condition.

Method

Participants

A total of 71 children aged 8 or 9 years participated in the study (M_{age} = 9.03 years, *SD* = 0.31; 28 boys). Participants were recruited from four primary schools in England. Ethical approval was obtained from the first author's institution, and informed parental consent was received for all participants. All children had normal or corrected-to-normal vision, and teachers confirmed an absence of learning or neurological disabilities. All children were monolingual native English speakers.

Materials and procedure

Background measures

Children completed background measures in one session prior to the word learning task. All were standardized assessments and were administered according to test manual instructions. Nonverbal abilities were measured using the Raven's Coloured Progressive Matrices (CPM; Rust, 2008), a pattern completion task in which participants solve visual diagrammatic puzzles using analogies or inferences (split-half reliability reported in the test manual = .97). Oral language abilities were assessed using the British Picture Vocabulary Scale–Third Edition (BPVS; Dunn, Dunn, & National Foundation for Educational Research, 2009), a receptive vocabulary measure in which children need to choose the correct picture for a given word among four alternatives, and the Understanding of Spoken Paragraphs (USP) subtest of the Clinical Evaluation of Language Fundamentals–Fourth Edition (CELF-4; Semel, Wiig, & Secord, 2006), a test in which children listen to several short passages and answer questions about them. This test assesses both oral language comprehension and sustained oral attention (test–retest reliability reported in the test manual = .80). Nonword reading was assessed using the Phonemic Decoding Efficiency subtest of the Test of Word and Reading Efficiency (TOWRE; Torgesen, Wagner, & Rashotte, 1999), in which children read as many nonwords as they can in 45 s (test–retest reliability reported in the test manual = .90), and word reading was assessed using the Single Word Reading Test (SWRT 6–16; Foster, 2007), which is an untimed word reading task with words in sets of increasing difficulty. The York Assessment of Reading for Comprehension (YARC; Snowling et al., 2009) was used to assess text reading accuracy and reading comprehension. In this task, children read two passages aloud and reading errors are noted. Following reading, they answer comprehension questions indexing knowledge of literal content and inferential processes (parallel form reliability reported in the test manual for reading accuracy: all *r*s > .70; Cronbach's alpha for reading comprehension scores from two passages: all αs > .70).

Word learning task

Design. Story presentation modality was manipulated between participants such that children were assigned to either the listening, reading, or combined condition in order to form three comparable groups matched on key background measures. The groups did not differ on gender, $\chi^2(2) = 0.09$, $p = .957$, reading, oral language, or nonverbal abilities (see Table 1). Details of the background measures used to match groups are included above. Children from each school were equally distributed across conditions. The presence of a definition was manipulated within participants, with all children encountering half of the words with a definition and half without a definition.

Table 1
Performance on background measures by story presentation modality.

Variable	Listening group (n = 24; 9 boys)		Reading group (n = 23; 9 boys)		Combined group (n = 24; 10 boys)		Difference between groups	
	Mean (*SD*)	Range	Mean (*SD*)	Range	Mean (*SD*)	Range	*F*	*p*
Age (years)[a]	9.06 (0.28)	8.58–9.50	8.96 (0.31)	8.50–9.50	9.07 (0.34)	8.41–9.50	2.05	.359
TOWRE PDE								
Raw score	34.38 (10.19)	9–51	36.04 (8.05)	23–53	35.13 (11.81)	13–59	0.16	.854
Standard score	101.46 (12.34)	71–121	104.96 (10.32)	86–126	103.17 (15.66)	73–145	0.43	.655
SWRT								
Raw score	43.63 (9.85)	17–57	43.87 (8.29)	23–55	42.83 (7.91)	27–54	0.09	.913
Standard score[a]	105.96 (15.22)	74–130	106.61 (13.89)	75–130	104.21 (13.69)	82–127	0.47	.792
YARC accuracy								
Ability score[a]	53.13 (9.13)	33–71	54.00 (8.47)	40–74	54.00 (7.81)	41–69	0.13	.938
Standard score	101.29 (11.06)	82–123	103.00 (11.44)	82–128	102.17 (10.28)	87–127	0.14	.867
YARC comprehension								
Ability score	58.63 (6.79)	37–68	59.78 (6.41)	46–77	59.33 (5.62)	48–68	0.20	.817
Standard score	104.08 (8.03)	83–118	105.78 (8.58)	87–128	104.63 (7.09)	90–118	0.28	.756
BPVS								
Raw score	121.96 (12.66)	94–144	120.87 (11.91)	94–146	122.58 (12.16)	94–143	0.11	.889
Standard score	96.71 (13.47)	79–119	97.74 (13.00)	70–122	97.46 (12.76)	70–115	0.04	.962
USP CELF								
Raw score[a]	11.33 (1.63)	7–14	11.43 (1.95)	8–15	11.38 (2.34)	6–15	0.14	.931
Scaled score[a]	10.25 (1.62)	7–13	10.09 (2.15)	7–14	10.38 (2.29)	6–14	0.48	.788
CPM								
Raw score	30.04 (3.46)	23–35	28.35 (3.83)	20–34	28.78 (3.41)	22–33	1.43	.246
Standard score[a]	105.00 (13.99)	80–125	100.65 (15.47)	75–130	101.25 (12.70)	80–125	1.33	.515

Note. TOWRE PDE = Phonemic Decoding Efficiency subtest of the Test of Word and Reading Efficiency; SWRT = Single Word Reading Test included in the York Assessment of Reading for Comprehension (YARC) protocol; YARC accuracy = reading accuracy of passages collected as part of the YARC; YARC comprehension = score associated with the reading comprehension of passages collected as part of the YARC; BPVS = score for the British Picture Vocabulary Scale; USP CELF = score for the Understanding of Spoken Paragraphs subtest of the Clinical Evaluation of Language Fundamentals; CPM = score obtained in the Raven's Coloured Progressive Matrices.

[a] The Kruskal–Wallis test is reported due to the measure being non-normally distributed in at least one of the groups.

Stimuli. See Appendix A for stimuli.

Words. Target items were 8 low-frequency English concrete nouns (e.g., destrier, hauberk), each belonging to one of eight categories (e.g., animal, object). From an initial set of 36 words related to historical periods (Roman Empire, Norman Period, and British Empire in India), we selected 8 words from the Norman Period and embedded these target items into a meaningful narrative. In addition, 8 control words were selected from other historical periods and were included in the pretest and all posttests to control for prior knowledge (see below for details of these tasks). The pretest confirmed that knowledge of target and control words was negligible (see below). Given that control words were not presented in the story, it was expected that any difference in performance between target and control words could be ascribed to story presentation. Target and control lists were composed of words of the same category, matched for length and frequency using the SUBTLEX–UK database (Van Heuven, Mandera, Keuleers, & Brysbaert, 2014) (Mann–Whitney tests, all *ps* > .200). Target and control words were also matched on pilot data collected for a previous study showing the proportion of adults who spelled and pronounced each word correctly and the proportion of children who correctly categorized each word at its first encounter (Mann–Whitney tests, all *ps* > .200).

Definitions. The definition for each target and control word comprised information about the word's subcategory and a further phrase to specify it. For example, for *destrier,* the definition "a horse used for fighting" comprises both the subcategory information "a horse" and the specific characteristic "used for fighting." Words were divided into two lists of four items (Word Lists A and B) for counterbalancing purposes. The words in the two lists were paired for length, frequency (Van Heuven et al., 2014), the aforementioned results from pilot studies, the length of the definition, and the distance

between the first and second mentions of the word in the story. Mann–Whitney tests revealed no difference between the words in the two lists on these measures (all *ps* > .200).

Stories. A story set during the Norman Period was written for this study to include all target words. Each target word was repeated three times in the story. Contextual references to the meaning of each word were included to ensure that the children could learn the meaning of the words from context alone.

Two different versions of the story were created so that the definitions of either Word List A or Word List B were included as part of the text following the first mention of the word. The two versions were matched for length (1346 and 1347 words, respectively), Flesh Reading Ease (84.3% and 84.5%, respectively) and Flesh–Kinkaid grade level (5.2). Pilot data ensured that the stories were written at an appropriate level for children in this age range. Recordings of a female native English speaker reading the stories were created, and booklets of seven pages that contained only the text of the stories, written in Calibri 14-point font, were prepared. Pilot studies established how long, on average, children took to read the stories independently (M = 548 s, SD = 172); the recordings of the adult storyteller were controlled to match this (M = 507 s, SD = 2).

Procedure

Fig. 1 summarizes the study procedure. Assessments were completed in a quiet room within schools. Prior to the word learning task, children completed the word knowledge pretest and were administered the background measures. The word learning task was completed in two sessions lasting between 30 min and 1 h. In the first session, children were exposed to the story for the first time and completed a story comprehension task. In the second session, they were exposed to the story a second time and completed the same story comprehension task as well as the phonological, orthographic, and semantic posttests. The first and second sessions were completed 1 week apart. The story comprehension task and all posttests were delivered through a laptop computer using E-Prime software (Schneider, Eschman, & Zuccolotto, 2002).

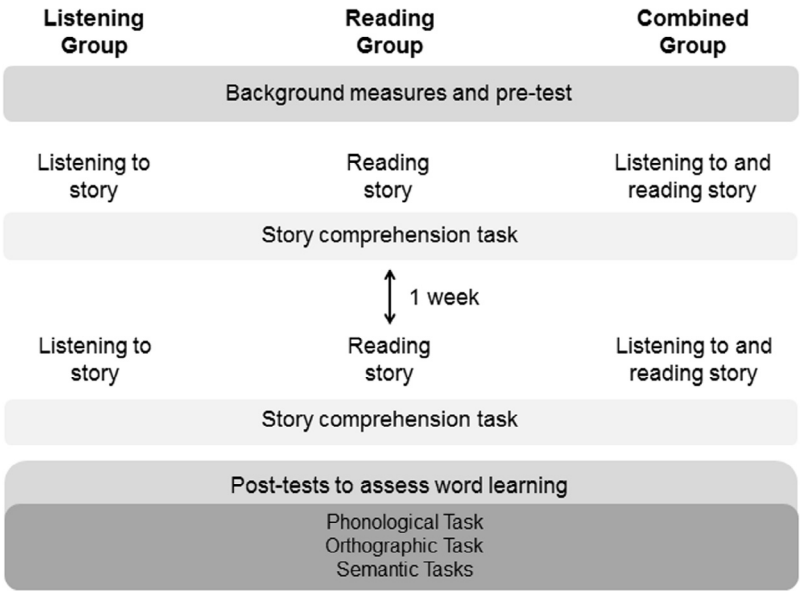

Fig. 1. Overview of study procedure for each group. Session 1: Collection of background measures. Session 2: Story presentation and story comprehension task. Session 3 (1 week later): second story presentation and completion of posttests assessing learning.

The phonological and orthographic posttests were completed first, in counterbalanced order, followed by the semantic posttests. Given that the semantic posttests involved presenting the novel words in spoken and written forms, these were completed last to avoid any contamination from these to the phonological and orthographic tasks. We did not expect that there would be cross-contamination between orthographic and phonological posttests and the semantic posttest because the former provided no semantic information. The semantic posttests were multiple-choice format and presented in a fixed order, with the category recognition task first, followed by the subcategory recognition task and finally the definition recognition task. This order minimized any impact of previous semantic tasks on later semantic tasks. For each posttest, on-screen instructions followed by four practice trials ensured that children understood the demands of the task. Target and control words were presented with item order randomized, and accuracy was recorded for each trial.

Word knowledge pretest

Children were asked to define the target and control words. Here, 3 participants showed preexisting knowledge of one target word (2 children knew *pottage* and 1 child knew *motte*); the remaining 68 children had no knowledge of any target words. In addition, 3 children showed preexisting knowledge of one control word (2 children knew *catacomb* and 1 child knew *verandah*); again, 68 children had no knowledge of any control words. Thus, preexisting knowledge of control and target words was similarly scarce. Analyses that excluded these participants yielded the same pattern of results as those reported. Thus, all participants were retained in the analysis.

Story exposure

Participants were asked to listen to (listening group), read silently at their own pace (reading group), or listen to and read (combined group) the story, after which they were told they would be assessed on their comprehension of the story. No mention was made of the presence of the target words. Children were presented with a version of the story that contained definitions for either Word List A or Word List B. The oral presentation of the story was delivered through headphones connected to a laptop with a blank screen. For the written presentation, children read the story from a booklet. In the combined condition, children received both presentations simultaneously.

Story comprehension task

After each story exposure, children were asked four multiple-choice questions to ensure that they had paid attention to the story (e.g., "Fred wanted to reach the king's castle. How long did he think the journey would last?"; correct response: *A month;* foils: *A day/A week/One hour*). Each question, along with an array of four potential answers, was presented both orally and visually via a laptop. Pilot data provided by 12 children of the same age as participants confirmed that, without exposure to the story, children were unlikely to guess the answers to questions at above chance levels.

Phonological posttest

In this task, two dinosaurs were presented sequentially on a computer screen: one providing the correct spoken form of a word and the other providing an incorrect word form (distractor). Pilot data collected for an earlier study were used to generate distractors for this task. Adults were asked to pronounce each written target and control word, and the most frequent mispronunciations were used as distractors for most words. If no mispronunciations were produced (for 4 words), alternative pronunciations were created. For example, the distractor for "hauberk" (hɔːbək, first vowel as in *h<u>or</u>n*) was "hɑːbək" (first vowel as in *h<u>ear</u>t*). Children were instructed to choose the dinosaur that "said the word best" by pressing the corresponding button on the keyboard. The dinosaurs appeared in turn for 2 s in an alternating loop until an answer was provided. The association between the two dinosaurs and the correct answer was randomized. Both target and control words were presented in this task (16 items). Test–retest reliability for these items was obtained from pilot data collected from a separate group of 57 children of the same age (8–9 years) who completed the tasks twice, with a 1-week interval between test sessions. Pilot data were available for 13 items used in the current study (all 8 target

words and 5 control words). Test–retest reliability for these items was $r(57) = .71$. See Appendix B for stimuli.

Orthographic posttest

As in the phonological task, children saw two dinosaurs sequentially, this time accompanied by a letter string, and were asked to choose the correct spelling of the given target or control word (by choosing the dinosaur who spelled the word best). The dinosaurs, alongside their spelling option, appeared in turn for 3.5 s in an alternating loop until an answer was provided. Pilot data provided by adults in an earlier study were used to generate distractor spellings. The most common misspellings of orally presented words were used; when no misspelling was produced (for 2 words), alternative spellings were created. For example, the distractor for "hauberk" was "horberk." The association between the two dinosaurs and the correct answer was randomized. Both target and control words were presented in this task (16 items). As for the phonological task, test–retest reliability was computed from pilot data available for 13 of the items: $r(57) = .57$. Although this value highlights a less than excellent relationship, we deemed the measure to be sufficiently reliable because an amount of variability is to be expected in representations of newly learned words over time. See Appendix B for stimuli.

Semantic posttests

The three semantic subtests followed the same format. In each task, the children were asked to choose the correct alternative from an array of four choices pressing the corresponding button on the keyboard. First, a category recognition task assessed recognition of the category of the new word (e.g., clothing) among three of the categories of other target words (e.g., part of a house, job, animal). At the beginning of each trial, a target or control word was presented in spoken and written forms; the written form appeared at the top of the screen. The alternative category labels appeared one at a time both orally and in written form in randomized positions on the screen (see Fig. 2). The second and third tasks followed the same procedure. The second task assessed recognition of the subcategory of the word (e.g., for an item of clothing, the four alternatives were different kinds of clothing), and the third task assessed recognition of the word's definition when presented with three distracter definitions (each differing from the correct choice by one feature). The definitions were not identical to those presented in the stories but were rephrased. Both target and control words were presented in all three semantic posttests (16 items). As for the other posttests, test–retest reliability for category recognition was computed from pilot data available for 13 of the items: $r(57) = .76$. Table 8 lists the correct responses for each semantic task.

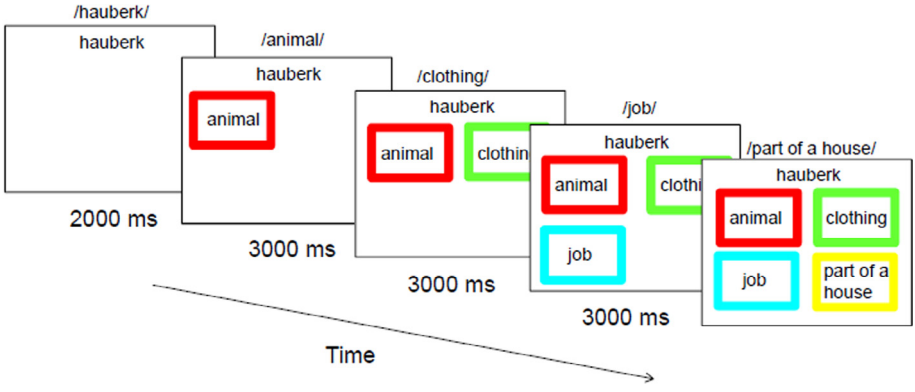

Fig. 2. Event sequence of a given trial of the semantic task (category recognition). The word was presented both orally and written, and then each of the four alternative category labels was presented both orally and written. The colored boxes around the four alternatives (red, green, blue and yellow) cued the appropriate response buttons on the keyboard. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.).

Results

Story comprehension

A story comprehension task was used to confirm that children paid attention to the story. Participants completed it twice, once after each story exposure, giving a maximum score of 8 and chance-level performance of 2. One-sample Wilcoxon signed-rank tests indicated that all groups performed better than chance on this task (see Table 2). A Mann–Whitney test showed that children in the combined group outperformed children in the reading group ($U = 140.50$, $p = .003$) but not children in the listening group ($U = 253.50$, $p = .461$), whereas there was a nonsignificant trend for the listening group to outperform the reading group ($U = 192.00$, $p = .069$). This suggests that hearing the story promoted comprehension.

Approach to data analysis

For each posttest, two sets of analyses were carried out. The first set (see Table 3; see also Table 6 below) compared the recognition of target and control words and compared each of these with chance (50% for phonological and orthographic tasks and 25% for semantic tasks). The second set of analyses used a mixed-effects modeling approach to explore our hypotheses relating to (a) presentation modality (listening, reading, or combined group), (b) definitions, and (c) individual differences within each of the three experimental tasks in turn. Because the groups were matched on age and all background measures, we included in the analyses only measures for which specific hypotheses were considered: vocabulary knowledge (BPVS raw score) and a composite reading accuracy score in all analyses, and reading comprehension (YARC reading comprehension ability score) in semantic task analyses. The composite reading accuracy score was formed by merging TOWRE, SWRT, and YARC text reading accuracy scores into a factor using the regression method ($M = 0.00$, $SD = 1.00$, range = −2.89 to 2.24). The three groups did not differ significantly on this measure, $H(2) = 0.02$, $p = .992$, and the creation of this factor was supported by high correlations between its constituent measures (all rs > .80, all ps < .001). See Appendix C for correlations between background measures.

Because the data collected on each trial were binomial (a child could choose either the correct or incorrect alternative, obtaining a score of 1 or 0), mixed-effects models were conducted using generalized linear mixed models for binomial data (Jaeger, 2008), using the function "glmer" from the package *lme4* (Bates, Maechler M., & Walker S., 2014), the function "mixed" from the package *afex* (Singmann, Bolker, Westfall, & Aust, 2016), and the function "lsmeans" from the package *lsmeans* (Lenth, 2016), computed with the software R (R Core Team., 2014). Each of the 71 children provided eight responses to target words on each task; this score was the dependent variable in each analysis. For each dependent variable, an initial model included a maximal random effects structure that

Table 2
Performance in the story comprehension task.

	Mean (*SD*)	Median	Range	Chance comparison	
				W	*p*
Listening group					
First session	2.88 (1.12)	3.00	1–4		
Second session	3.37 (0.92)	4.00	1–4		
Total	6.25 (1.81)	7.00	2–8	276.00	<.001
Reading group					
First session	2.48 (0.99)	2.00	1–4		
Second session	2.87 (2.87)	3.00	1–4		
Total	5.35 (1.84)	5.00	2–8	231.00	<.001
Combined group					
First session	3.21 (0.68)	3.00	1–4		
Second session	3.63 (0.49)	4.00	3–4		
Total	6.83 (1.04)	7.00	4–8	300.00	<.001

Table 3
Performance in the phonological and orthographic posttests.

Task	Group	Items	Mean (*SD*)	Range	Chance Comparison		Target vs. Control	
					W	*p*	*T*	*p*
Phonological task	Listening group	Target words	6.42 (1.10)	4–8	253.00	.001	0.00	<.001
		Control words	4.58 (1.50)	1–7	124.00	.087		
	Reading group[b]	Target words[a]	5.43 (1.53)	1–8	4.49	<.001	3.74	.001
		Control words[a]	3.91 (1.41)	1–7	-0.30	.770		
	Combined group	Target words	6.67 (1.69)	1–8	244.00	<.001	29.00	.002
		Control words	5.29 (1.30)	3–8	165.00	<.001		
Orthographic task	Listening group	Target words	4.54 (1.44)	2–8	137.50	.076	100.50	.598
		Control words[a]	4.37 (1.31)	2–7	1.40	.175		
	Reading group	Target words[a]	6.00 (1.38)	3–8	6.94	<.001	6.00	<.001
		Control words	4.39 (1.16)	2–6	85.00	.143		
	Combined group	Target words[a]	6.21 (1.35)	3–8	8.01	<.001	3.00	<.001
		Control words	4.13 (1.33)	1–6	97.00	.603		

[a] One-sample t-test are reported in place of one-sample Wilcoxon signed-rank test – the distribution of the Sample is normal.
[b] Paired t-test is reported in place of Wilcoxon signed-rank test – the distribution of both control and target words is normal.

captured our experimental design (Barr, Levy, Scheepers, & Tily, 2013). This entailed the random intercept terms for both participants and items and the random slopes terms for participants and items that relate to our repeated-measures manipulation: presence of a definition. However, models including random slopes were prone to nonconvergence; therefore, the simpler and convergent models are reported (Bates, Kliegl, Vasishth, & Baayen, 2015). We then compared these "empty" models (using pairwise likelihood ratio test comparisons; Barr et al., 2013) with models that also included performance on control words as a control variable and the hypothesized fixed effects: group (combined, listening, or reading), presence of definition (definition present or definition absent), and specific background measures. Scores on control word trials were included in models to control for general task effects, such as the ability to recognize word-like phonological forms, irrespective of any in-task learning. Further analyses were performed to explore whether results differed for the two sets of target words by entering word set as a further fixed factor. The pattern of results was identical, and set was not a significant predictor within the models; thus, these models have not been reported.

All continuous factors were centered around the mean for analysis. Hypothesized interactions were included one at a time in the model with all fixed effects and were retained only if significant. The interactions between group and the background measures were separately introduced to test whether any background measure had a differential effect on performance in each task, depending on presentation modality. Estimates of fixed effects and interactions for the final models are reported in Tables 4, 5, and 7 below.

Phonological task

The mean number of phonological forms correctly recognized by the children in each group is presented in Table 3 along with analyses comparing target and control word performance with each other and with chance. Target word performance was greater than control word performance and was above chance for all groups. Control word performance was also greater than chance for the combined group, but this was not the case for the listening and reading groups.

The fixed-factor model for the phonological task significantly improved fit compared with the empty model, $\chi^2(6) = 48.53$, $p < .001$. Interaction terms did not significantly improve model fit. The final model (Table 4) indicates reading accuracy and control word scores as significant predictors of performance on the phonological task; phonological learning was greater for better readers and those better able to identify the phonological forms of control words. Most important, phonological learning did not differ across groups.

Table 4
Generalized linear mixed model for performance on the phonological task.

Factor	Estimate	SE	*z* values		χ^2	
			z Value	*p*	χ^2	*p*
Intercept	.76	.21	3.57	<.001		
Group					3.45	.180
Combined vs. Listening	−.09	.24	0.37	.928		
Combined vs. Reading	.31	.24	1.26	.416		
Listening vs. Reading	.39	.23	1.74	.196		
Presence of definition	.10	.19	0.55	.582	0.32	.570
Reading accuracy	.21	.10	2.15	.031	4.85	.030
Vocabulary	.12	.10	1.19	.233	1.79	.180
Control words	.56	.10	5.47	<.001	31.19	<.001

Orthographic task

The mean number of orthographic forms correctly recognized by the children in each group is presented in along with analyses comparing target and control word performance with each other and with chance. In the Table 3orthographic task, target word performance was significantly greater than control word performance and was above chance for the reading and combined groups. For the listening group, target word performance was not significantly greater than control word performance or above chance. Control word performance was not above chance for any group.

The fixed-factor model for the orthographic task significantly improved fit compared with the empty model, $\chi^2(6) = 31.53$, $p < .001$. Interaction terms did not significantly improve model fit. The final model (Table 5) indicates group and reading accuracy as significant predictors of orthographic performance, with children in the combined and reading groups showing better performance than children in the listening group (who did not show significant orthographic learning) and greater orthographic learning associated with greater reading accuracy.

Semantic tasks

The mean number of words correctly recognized by each group in each semantic task is presented in Table 6 along with analyses comparing target and control word performance with each other and with chance. In all three semantic tasks, all groups performed significantly above chance on target words and significantly better on target word trials than on control word trials. Control word performance was not significantly above chance in any task for group. The final model for each semantic task is presented in Table 7.

Table 5
Generalized linear mixed model for performance on the orthographic task.

Factor	Estimate	SE	*z* values		χ^2	
			z Value	*p*	χ^2	*p*
Intercept	1.42	.30	4.77	<.001		
Group					20.14	<.001
Combined vs. Listening	1.03	.25	4.11	<.001		
Combined vs. Reading	.15	.26	0.55	.842		
Listening vs. Reading	−.88	.25	-3.58	.001		
Presence of definition	−.08	.20	-0.40	.689	0.16	.690
Reading accuracy	.36	.11	3.22	.001	10.42	.001
Vocabulary	.06	.11	0.55	.581	0.32	.570
Control words	−.19	.11	-1.73	.085	3.00	.080

Table 6
Performance in the semantic posttests: Numbers of words correctly recognized by the three groups.

		Category recognition	Chance comparison		Target vs. Control		Subcategory recognition	Chance comparison		Target vs. Control		Definition recognition	Chance comparison		Target vs. Control	
		Mean (*SD*)	*W*	*p*	*T*	*p*	Mean (*SD*)	*W*	*p*	*T*	*p*	Mean (*SD*)	*W*	*p*	*T*	*p*
Listening group	Target words	3.67 (1.69)	4.85[a]	<.001	25.50	.003	4.63 (1.56)	210.00	<.001	10.50	<.001	4.50 (1.77)	267.00	<.001	10.00	<.001
	Control words	2.42 (1.25)	130.00	.129			2.54 (1.22)	116.00	.054			1.75 (1.26)	58.50	.379		
Reading group	Target words	3.65 (1.37)	203.00	<.001	13.50	<.001	4.35 (1.43)	7.85[a]	<.001	0.00	<.001	4.13 (1.89)	206.50	<.001	0.00	<.001
	Control words	1.70 (1.22)	55.50	.302			2.13 (1.10)	98.00	.559			1.96 (0.93)	42.50	.552		
Combined group	Target words	4.71 (1.60)	8.29[a]	<.001	7.50	<.001	4.96 (1.57)	9.21[a]	<.001	15.00	<.001	5.00 (1.47)	9.97[a]	<.001	0.00	<.001
	Control words	1.92 (1.21)	67.00	.635			2.42 (1.10)	89.00	.084			1.83 (1.09)	55.00	.479		

[a] One-sample *t* tests are reported in place of Wilcoxon signed-rank test.

Table 7
Generalized linear mixed model for performance in the semantic posttests.

Factor	Estimate	SE	*z* values		χ^2	
			z Value	*p*	χ^2	*p*
Category recognition						
Intercept	.64	.27	2.35	.019		
Group					9.97	.007
Combined vs. Listening	.67	.23	2.93	.009		
Combined vs. Reading	.55	.23	2.44	.039		
Listening vs. Reading	−.11	.23	-0.48	.879		
Presence of definition	−.42	.18	-2.28	.023	5.05	.020
Reading accuracy	.33	.10	3.24	.001	10.03	.002
Reading comprehension	.07	.11	0.65	.519	0.16	.690
Vocabulary	.40	.11	3.59	<.001	13.00	<.001
Control words	.18	.09	1.98	.048	3.69	.050
Subcategory recognition						
Intercept	.38	.36	1.04	.297		
Group					1.83	.400
Combined vs. Listening	.19	.24	0.80	.705		
Combined vs. Reading	.33	.25	1.36	.364		
Listening vs. Reading	.14	.25	0.57	.837		
Presence of definition	.45	.19	2.42	.016	5.94	.010
Reading accuracy	.14	.11	1.28	.201	1.61	.200
Reading comprehension	.09	.12	0.74	.463	0.48	.490
Vocabulary	.24	.12	1.96	.050	3.72	.050
Control words	.10	.11	0.87	.385	0.69	.400
Definition recognition						
Intercept	.34	.26	1.31	.191		
Group					3.84	.150
Combined vs. Listening	.22	.23	0.94	.614		
Combined vs. Reading	.46	.23	1.96	.121		
Listening vs. Reading	.24	.24	1.01	.572		
Presence of definition	.45	.18	2.45	.014	6.04	.010
Reading accuracy	.10	.11	0.95	.341	0.71	.400
Reading comprehension	.06	.20	0.28	.780	0.06	.800
Vocabulary	.34	.12	2.95	.003	8.45	.004
Control words	.18	.10	1.81	.070	3.32	.070
Group * Reading Comprehension					8.60	.010
Listening vs. Combined	.46	.27	1.69	.090		
Reading vs. Combined	−.30	.26	−1.18	.237		
Reading vs. Listening	−.76	.25	−3.01	.003		

Category recognition

The fixed-factor model significantly improved fit compared with the empty model, $\chi^2(7) = 55.34$, $p < .001$. Interaction terms did not significantly improve model fit. Group, presence of definition, reading accuracy, vocabulary, and control word scores were significant predictors in the final model. The combined group performed better than the other two groups, and greater reading accuracy, oral vocabulary knowledge, and performance on control words was associated with better performance. Category recognition was surprisingly better for words presented without definitions than for those presented with definitions.

Subcategory recognition

The fixed-factor model significantly improved fit compared with the empty model, $\chi^2(7) = 21.15$, $p < .001$. Interaction terms did not significantly improve model fit. Presence of definition and vocabulary level significantly predicted performance, with the presence of a definition and greater existing oral vocabulary knowledge associated with better performance. No group effect was evident; the three groups performed similarly on this task.

Definition recognition

The fixed-factor model significantly improved fit compared with the empty model, $\chi^2(7) = 42.61$, $p < .001$. Addition of the group by reading comprehension interaction also improved model fit. As for subcategory recognition, the presence of a definition and oral vocabulary knowledge were significant predictors, but group was not a significant predictor.

To further explore the group by reading comprehension interaction in the definition recognition task, a separate model was computed for each group. For the combined group, the presence of a definition predicted performance, $\beta = .97$, $\chi^2(1) = 9.63$, $p = .002$. The performance of the listening group was positively influenced by vocabulary, $\beta = .42$, $\chi^2(1) = 4.61$, $p = .030$, and reading comprehension, $\beta = .47$, $\chi^2(1) = 5.32$, $p = .020$, whereas the performance of the reading group was positively influenced by vocabulary, $\beta = .42$, $\chi^2(1) = 4.39$, $p = .040$, and control word scores, $\beta = .79$, $\chi^2(1) = 10.60$, $p = .003$. These supplementary analyses suggest that the group by reading comprehension interaction reflects a positive association between reading comprehension and performance for the listening group but not for the other two groups. In addition, the presence of definitions may have particularly supported later definition recognition by children in the combined group, whereas existing vocabulary knowledge particularly influenced performance in the listening and reading groups. However, because the group by definition and group by vocabulary interactions were not significant, no strong conclusions can be drawn in relation to these findings.

Discussion

This study investigated children's incidental word learning from stories, comparing listening, reading, and combined conditions for the first time. Children learned information about the phonology, orthography, and semantics of new words, but the extent and nature of their learning depended on the presentation modality. Learning was also modulated by the presence of a definition and by children's existing abilities. The influence of presentation modality, presence of a definition, and individual differences are discussed in turn.

In relation to presentation modality, the orthographic facilitation reported in previous studies (Rosenthal & Ehri, 2008) motivated our hypothesis that the combined group would outperform the other two groups on the phonological task. However, in our paradigm, where children learned new words incidentally rather than being taught them, children learned the phonological forms of the new words with equal proficiency across conditions. Thus, children who read the story appear to have formed a phonological representation of the new words that not only was better than chance and better than for control words but also was equivalent to the learning shown by children who had been directly provided with the phonological form. This is in line with Share's (1995) self-teaching hypothesis, which states that phonological recoding occurs while reading. Children were able to use their knowledge of the relationship between orthography and phonology to learn the phonological form of words that they saw in written format only.

To be certain that children's phonological representations were equivalent across the three conditions, however, we must be confident that our phonological task provided a robust measure of phonological learning. Three issues are worthy of mention in relation to this point. First, performance on this task involved distinguishing between targets and plausible foils, which were generated, where possible, from adult mispronunciations of the written forms. Consequently, it is possible that the task probed abilities other than children's in-task learning such as their general sensitivity to an oral form's word-likeness. This idea is supported by the finding that control word scores significantly predicted target word scores on this task. To account for such general effects, scores on control word trials, therefore, were included in all analytical models, enabling us to have confidence that our results reflect children's phonological learning within the task. A second consideration is whether the use of plausible foils made the task particularly challenging for children in the reading group, who were not provided with a phonological form and who, therefore, may have generated an alternative form while reading that was aligned more closely with the foil than with the target. However, this does not seem to have been the case; the reading group performed just as well as the other groups on the phonological task. Finally, it is possible that the recognition task was insufficiently sensitive to

detect subtle differences in the quality of children's phonological representations across conditions due to either the nature of the task itself or the number of items tested. Had we used a production task, for example (cf. Rosenthal & Ehri, 2008), group differences might have been detected. However, an identical recognition task format with the same numbers of items elicited group differences on the orthographic task in the current study (see below), and similar tasks and trial numbers have been found to be sufficiently sensitive in previous studies (e.g., Ricketts et al., 2009; Ricketts et al., 2011; Rosenthal & Ehri, 2008; Rosenthal & Ehri, 2011). Nevertheless, future studies might seek to corroborate our conclusions using alternative measures of phonological learning.

In contrast to the findings relating to phonological learning, the three groups did not show equivalent orthographic learning. Here, children in the combined and reading groups outperformed those in the listening group, whose performance was at chance. This indicates that a word's orthographic form is not automatically extrapolated from its phonology; rather, the presentation of written text prompts orthographic learning. Contrary to previous findings (Rosenthal & Ehri, 2011), the additional presentation of the oral form did not enhance orthographic learning relative to the presentation of the written form alone. As for the phonological task, it is possible that a more sensitive measure of orthographic knowledge might have revealed a difference in the orthographic representation of the two reading groups. Nevertheless, there is no evidence in our study for phonological facilitation of orthographic learning. Given the relatively low reliability for the orthographic task, these findings warrant replication in future studies.

In summary, the results of the phonological and orthographic recognition tasks suggest that listening to stories promotes learning of new phonological but not orthographic forms. Reading, in contrast, appears to support the acquisition of both phonological and orthographic information about new words, presumably creating higher quality lexical representations (cf. Perfetti & Hart, 2002). The asymmetry in our results resonates with observations that children tend to perform better in reading tasks that require them to produce an oral form from a written one than in spelling tasks that require the reverse (Cossu, Gugliotta, & Marshall, 1995).

Turning to the semantic tasks, we explored whether semantic learning would be greater in the reading group than in the listening group (Brown et al., 2008) or vice versa (Suggate et al., 2013). Our results indicated no support for the superiority of either oral or written presentation. It appears that, at 8 or 9 years of age, children learn as much about the meanings of words from reading as they do from listening to stories. It was also hypothesized that children in the combined condition would acquire more semantic information about new words than children in the listening and reading conditions (Ricketts et al., 2009; Rosenthal & Ehri, 2008; Rosenthal & Ehri, 2011). This prediction was upheld, but only for the category recognition task, which required the abstraction of word knowledge to categorize the word correctly. In this task, we observed both an orthographic facilitation effect (i.e., better performance in the combined group than in the listening group) and a phonological facilitation effect (i.e., better performance in the combined group than in the reading group).

Before we consider why the benefit of the combined condition was found in only one semantic task, we first reflect on the causes of children's better performance in the combined condition. The lexical quality hypothesis (cf. Perfetti & Hart, 2002) suggests that words with higher quality representations are more easily retrieved from memory. This theoretical perspective focuses on existing lexical representations rather than their acquisition. Nonetheless, it is consistent with the proposal that the combined condition promotes the building of better specified representations, facilitating access to stored word knowledge at test. However, this framework would appear to predict consistent differences in the quality of the phonological, orthographic, and semantic representations formed in the combined and single modality presentations, as reported previously in studies finding orthographic facilitation effects (Ricketts et al., 2009; Rosenthal & Ehri, 2008) and phonological facilitation effects (Rosenthal & Ehri, 2011). Such consistent effects across tasks were not found in the current study. As discussed above, it is possible that the insensitivity of our phonological and orthographic measures masked subtle differences between the groups. Although we find no evidence for this, it remains possible that the phonological or orthographic forms acquired in the combined condition were of a higher quality than those in the other two conditions.

An alternative possibility is that the combined condition reduced cognitive load during word learning, freeing resources for comprehension and word meaning extraction (Mayer et al., 1999). Compared

with the combined group, the reading and listening groups were charged with additional processing demands at the point of encountering the new words—the reading group in the form of spontaneous phonological recoding (as evidenced by the children's performance on the phonological task) and the listening group due to the attentional demands associated with continuously monitoring the oral story presentation (without any "backup" support from the written text). Children in the combined group, therefore, may have had more resources available to allocate to processing the contextual support (including definitions) that immediately followed a new word, thereby encoding the word meanings better. This perspective would predict word form representations to be of similar quality across conditions but would predict representations of words' meanings to be better in the combined condition, the pattern found in the current study.

Although both approaches could, in theory, explain the greater semantic learning of the combined group, it remains to be discussed why this effect was seen in only one of our semantic tasks, the category recognition task. Two potential explanations occur to us. First, only the category recognition task required children to abstract category-level knowledge about the target words from the information provided in the story: the category label for each new word was never directly provided in the narrative. In contrast, recognizing the correct subcategory or definition in the other semantic tasks required participants to choose a subcategory or definition that was very similar to those provided in the story. Second, choosing the correct response on the subcategory and definition recognition subtests is likely to have been easier than choosing the correct form in the category recognition task. In the former tasks, only one of the four alternative response options had been encountered in the story (e.g., among the four clothing options offered in the definition task for the word *hauberk*, only *a soldier's shirt made of chain mail* had been mentioned in the story, making the other alternatives less likely regardless of any learning of the label for this item). In contrast, all four of the alternatives in the category recognition task correctly described one of the new words presented in the story (i.e., to recognize *hauberk* as a piece of *clothing* in the category recognition task, children needed to know that this new word did not identify a new *animal, job,* or *part of a house,* each of which had been encountered in the story). These factors could have made the category recognition task the most challenging measure, and therefore the most sensitive measure, of children's semantic learning. By freeing resources, the combined condition might have especially facilitated this task due to the level of abstraction and/or precision of the mapping required, a hypothesis that warrants further investigation.

The study also investigated the impact of definitions on word learning. It was hypothesized that the presence of an accompanying definition alongside each new word would foster semantic learning (Wilkinson & Houston-Price, 2013). Although definitions did not affect children's phonological or orthographic learning in this study, they facilitated the learning of subcategories and definitions of the words but hindered category recognition. These findings indicate that definitions help children to learn detailed information regarding new words but do not help to extract more general categorical information. This finding is perhaps not surprising considering that in the current study the definitions directly provided most of the information needed in the definition recognition task along with the subcategory (or a synonym of this) required in the subcategory recognition task. When definitions were provided, the level of abstraction required by these two tasks, therefore, was quite low. When definitions were not provided, children needed to extract the details of the word's semantics from the text, synthesizing various cues to construct a coherent representation of its meaning, in order to succeed on these tasks. In contrast, the abstraction required to identify words' categories, necessary to succeed on the category recognition task, would be similar whether or not a definition was provided because this information was not supplied within definitions. When provided with a definition, children appear to have learned the specific information in the definition at the expense of abstracting a more general representation of the word's meaning.

We also explored the effects of children's reading and language abilities on their learning in different presentation conditions. Reading accuracy predicted performance in the phonological, orthographic, and category recognition tasks irrespective of presentation modality. Children with better baseline knowledge of orthography–phonology mappings were better able to learn new orthographic and phonological forms and extract semantic information (Ricketts et al., 2011; Rosenthal & Ehri, 2008). By facilitating the learning of new word forms, better decoding skills could potentially reduce

the cognitive load in a similar way to that discussed earlier in relation to the dual presentation modality, enabling the learning of more complex semantic information (category learning). As hypothesized, oral vocabulary level predicted performance in all three semantic tasks. It is likely that children with smaller vocabularies face a more difficult task when reading or listening to stories because their reduced familiarity with the vocabulary in the story increases the size of the word learning challenge (Shefelbine, 1990). In addition, children with larger vocabularies may employ better word learning strategies (Cain et al., 2004); for example, they may be more proficient at linking new words with contextual information (McKeown, 1985). Written text comprehension ability was also positively related to definition recognition, but only in the listening group. This finding is surprising, because children in the listening group were not required to comprehend written text. It is possible that a third factor—an unmeasured one—is involved in both reading comprehension and word learning while listening to stories, such as the ability to build a representation of the incoming story or other discourse online without the need to revisit the text (e.g., by rereading).

Despite the broad variation in children's abilities, it is worth noting that background measures did not interact with the impact of presentation modality on category recognition. It is particularly interesting that, irrespective of reading ability, children's category learning benefited from combined oral and written presentation. Thus, even good readers can benefit from oral presentation while reading, despite their ability to read efficiently by themselves. Similarly, poor readers can benefit from the presence of the written text alongside an oral presentation despite their poorer reading abilities. Given the growing use of e-books, these results are important in suggesting that listening to a narration of the story while reading enhances vocabulary acquisition. Having said that, given the narrow age range of the children in our study, it is possible that individual differences in ability might interact with presentation modality to affect learning in a younger or older population of school-aged children. Older, more efficient readers might, for example, learn equally well from reading and reading and listening, as adults do (Sydorenko, 2010), whereas less experienced readers might not be able to take advantage of a dual presentation modality compared with listening. Future studies that explore the relationship between the effect of presentation modality and individual differences in ability in younger and older readers would elucidate whether the pattern found in the current study changes with development.

Conclusion

This study shows that children aged 8 or 9 years are able to extract information about new words from a story and that the modality of the story presentation influences the nature and extent of their learning. Children learned the new words' phonology similarly well from the three presentation modalities (listening, reading, and combined conditions), but orthography was learned only when it was directly presented (reading and combined conditions). Children showed the strongest learning for words' semantic categories when words were presented both orally and in written form (combined modality).

The practical implications of these findings for the classroom are that children are able to learn information about the phonological forms, orthographic forms, and meanings of new words when they are listening to and/or reading stories. Importantly, both listening to and reading stories can support the learning of new phonological forms, whereas opportunities to see the new words written down are crucial for building representations of their orthographic forms. Finally, allowing children to hear stories while they are reading along may be optimal for learning, and especially for the extraction of semantic information, supporting teachers' practice of reading aloud in the classroom.

Acknowledgments

This research was supported by a University of Reading Research Studentship (Social Science) to the first author. The second author is supported by the Economic and Social Research Council (Grant ES/K008064/1). This study was presented at the 23rd annual meeting of the Society for the Scientific Study of Reading in July 2016. We thank all the children who participated, as well as the schools that participated in the study. We also thank Lotte Meteyard for her support with analyses.

Please note that data from this project are not archived due to the nature of the consent obtained from participants. All other enquiries about the project may be addressed to the corresponding author.

Appendix A. Stimuli for the word learning task.

Word	Definition within text	Contextual Reference 1 within text	Contextual Reference 2 within text	Correct category for category recognition	Correct subcategory for subcategory recognition	Correct definition for definition recognition
furrier	a hunter who sells animal's furs	experienced in hunting	working with furs all day	job	huntsman	someone who hunts animals to sell their fur
motte	hill built by men, on top of which a castle is built.	live on the … built by my ancestors	… on which the king's castle was built	building	man-made hill	a hill with a castle on top
palisade	barrier made of wooden stakes	a garden, surrounded by …..… surrounds my garden	wooden	part of a house	fence	a fence made of wood
pottage	thick liquid food that farmers usually eat	watery	farmer's …	food or drink	soup	soup eaten by farmers
wain	cart, pulled by an animal	we can use to transport the materials we need	the donkey was unable to heave it up the steep path	vehicle	wagon	a vehicle drawn by horses oxen or mules
destrier	horse used for fighting	horse we had been given	used to knightly fights	animal	horse	war horse
hauberk	piece of armour that covers the top of the body, made of metal chains	to protect my chest	made of rusty chains	clothing	shirt	a soldier's shirt made of chain mail
trencher	a flat dishmade ofwood	roasted pork on a …..…full of meat and pies	wooden trenchers	object	plate	a wooden plate

Appendix B. Word pronunciations and orthographic and phonological alternatives (following IPA transcription) for the orthographic and phonological tasks.

Set	Word	Orthographic foil	Word pronunciation	Phonological foil
Target words	motte	mot	mɒt	mɒtɪ
	furrier	fourrier	fʌrɪə	fʊrɪə
	hauberk	horberk	hɔːbək	hɑːbək
	destrier	destriar	dɛstrɪə	dɛstraɪə
	palisade	palisaid	palɪseɪd	palɪsɑːd
	wain	waine	weɪn	waɪən
	pottage	potage	pɒtɪdʒ	pɒtɑːdʒ
	trencher	trencha	trɛntʃə	trɪntʃə
Control words	catacomb	catacoomb	kætəkuːm	kætəkɒm
	augur	auger	ɔːgə	eɪgə
	dhoti	dottie	dəʊtɪ	dɔɪtɪ
	gavial	gaviel	geɪvɪəl	gævɪəl
	atrium	atriam	eɪtrɪəm	ætrɪəm
	palanquin	pelanquin	palənjkwɪn	palənjkwiːn
	toddy	toddie	tɒdɪ	tuːdɪ
	teapoy	teapoi	tiːpɔɪ	tepɔɪ

Note. Pronunciations and phonological foils are conveyed using the International Phonetic Alphabet (IPA).

Appendix C. Spearman correlation coefficients between the background measures.

	Reading accuracy factor score	BPVS raw score	YARC comprehension ability score	USP CELF raw score	CPM raw score
Reading accuracy factor score	–				
BPVS raw score	.16	–			
YARC comprehension ability score	.12	.46**,[a]	–		
USP CELF raw score	−.03	.34**	.36**	–	
CPM raw score	.38**	.39**,[a]	.21[a]	.37**	–

Note. YARC comprehension = score associated with the reading comprehension of passages collected as part of the York Assessment of Reading for Comprehension; BPVS = score for the British Picture Vocabulary Scale; USP CELF = Understanding of Spoken Paragraphs subtest of the Clinical Evaluation of Language Fundamentals; CPM = score obtained in the Raven's Coloured Progressive Matrices.

[a] Pearson correlations are reported; the measures are normally distributed.

** Correlation is significant at $p < .01$.

References

Akhtar, N. (2004). Contexts of early word learning. In O. G. Hall & S. R. Waxman (Eds.), *Weaving a lexicon* (pp. 485–507). Cambridge, MA: MIT Press.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278.

Bates, D., Maechler M., Bolker B., & Walker S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7. http://CRAN.R-project.org/package=lme4.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. Retrieved from http://arxiv.org/pdf/1506.04967.pdf.

Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language, 20*, 136–163.

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.

Cossu, G., Gugliotta, M., & Marshall, J. C. (1995). Acquisition of reading and written spelling in a transparent orthography: Two non-parallel processes? *Reading and Writing, 7*, 9–22.

Dickinson, D. K. (1984). First impressions: Children's knowledge of words gained from a single exposure. *Applied Psycholinguistics, 5*, 359–373.

Dunn, L. M., Dunn, D. M., & National Foundation for Educational Research (2009). *British Picture Vocabulary Scale* (3rd ed.). London: GL Assessment.

Ehri, L. C. (2014). Orthographic mapping in the acquisition of sight word reading, spelling memory, and vocabulary learning. *Scientific Studies of Reading, 18*, 5–21.

Elley, W. B. (1989). Vocabulary acquisition from listening to stories. *Reading Research Quarterly, 24*, 174–187.

Foster, H. (2007). *Single Word Reading Test 6–16*. London: GL Assessment.

Hartman, F. R. (1961). Single and multiple channel communication: A review of research and a proposed model. *Audiovisual Communication Review, 9*, 235–262.

Henderson, L., Devine, K., Weighall, A., & Gaskell, G. (2015). When the daffodat flew to the intergalactic zoo: Off-line consolidation is critical for word learning from stories. *Developmental Psychology, 51*, 406–417.

Henderson, L., Weighall, A., & Gaskell, G. (2013). Learning new vocabulary during childhood: Effects of semantic training on lexical consolidation and integration. *Journal of Experimental Child Psychology, 116*, 572–592.

Houston-Price, C., Howe, J. A., & Lintern, N. J. (2014). Once upon a time, there was a fabulous funambulist Ellipsis: What children learn about the "high-level" vocabulary they encounter while listening to stories. *Frontiers in Psychology, 5*. https://doi.org/10.3389/fpsyg.2014.00075.

Hu, C. F. (2008). Use orthography in L2 auditory word learning: Who benefits? *Reading and Writing, 21*, 823–841.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*, 434–446.

Justice, L. M., Meier, J., & Walpole, S. (2005). Learning new words from storybooks: An efficacy study with at-risk kindergartners. *Language, Speech, and Hearing Services in Schools, 36*, 17–32.

Lenth, R. V. (2016). Least-squares means: The R package *lsmeans*. *Journal of Statistical Software, 69*, 1–33.

Lin, L. C. (2014). Learning word meanings from teachers' repeated story read-aloud in EFL primary classrooms. *English Language Teaching, 7*(7), 68–81.

Mayer, R. E., Moreno, R., Boire, M., & Vagge, S. (1999). Maximizing constructivist learning from multimedia communications by minimizing cognitive load. *Journal of Educational Psychology, 91*, 638–643.

McKeown, M. G. (1985). The acquisition of word meaning from context by children of high and low ability. *Reading Research Quarterly, 20*, 482–496.

Menne, J. M., & Menne, J. W. (1972). The relative efficiency of bimodal presentation as an aid to learning. *Educational Technology Research and Development, 20*, 170–180.

Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology, 87*, 319–334.

Nagy, W. E. (1995). *On the role of context in first- and second-language vocabulary learning*. Champaign: University of Illinois at Urbana-Champaign, Center for the Study of Reading.

Nagy, W. E., Anderson, R. C., & Herman, P. A. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal, 24*, 237–270.

Penno, J. F., Wilkinson, I. A., & Moore, D. W. (2002). Vocabulary acquisition from teacher explanation and repeated listening to stories: Do they overcome the Matthew effect? *Journal of Educational Psychology, 94*, 23–33.

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

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. Vehoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189–213). Philadelphia: John Benjamins.

R Core Team. (2014). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org.

Ricketts, J., Bishop, D. V., & Nation, K. (2009). Orthographic facilitation in oral vocabulary acquisition. *Quarterly Journal of Experimental Psychology, 62*, 1948–1966.

Ricketts, J., Bishop, D. V., Pimperton, H., & Nation, K. (2011). The role of self-teaching in learning orthographic and semantic aspects of new words. *Scientific Studies of Reading, 15*, 47–70.

Ricketts, J., Dockrell, J. E., Patel, N., Charman, T., & Lindsay, G. (2015). Do children with specific language impairment and autism spectrum disorders benefit from the presence of orthography when learning new spoken words? *Journal of Experimental Child Psychology, 134*, 43–61.

Rosenthal, J., & Ehri, L. C. (2008). The mnemonic value of orthography for vocabulary learning. *Journal of Educational Psychology, 100*, 175–191.

Rosenthal, J., & Ehri, L. C. (2011). Pronouncing new words aloud during the silent reading of text enhances fifth graders' memory for vocabulary words and their spellings. *Reading and Writing, 24*, 921–950.
Rust, J. (2008). *Coloured Progressive Matrices and Chrichton vocabulary scale manual*. London: Pearson.
Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-Prime (version 2.0) [computer software and manual]. Pittsburgh, PA: Psychology Software Tools.
Semel, E., Wiig, E., & Secord, W. (2006). *Clinical Evaluation of Language Fundamentals-Fourth Edition (CELF-4 UK)*. London: Pearson.
Senechal, M., Thomas, E., & Monker, J. A. (1995). Individual differences in 4-year-old children's acquisition of vocabulary during storybook reading. *Journal of Educational Psychology, 87*, 218–229.
Shamir, A., Korat, O., & Fellah, R. (2012). Promoting vocabulary, phonological awareness and concept about print among children at risk for learning disability: Can e-books help? *Reading and Writing, 25*, 45–69.
Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition, 55*, 151–218.
Shefelbine, J. L. (1990). Student factors related to variability in learning word meanings from context. *Journal of Literacy Research, 22*, 71–97.
Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2016). *afex: Analysis of factorial experiments*. R package version 0.16-1. https://CRAN.R-project.org/package=afex.
Snowling, M. J., Stothard, S. E., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E., & Hulme, C. (2009). *York Assessment of Reading for Comprehension*. London: GL Assessment.
Suggate, S. P., Lenhard, W., Neudecker, E., & Schneider, W. (2013). Incidental vocabulary acquisition from stories: Second and fourth graders learn more from listening than reading. *First Language, 33*, 551–571.
Sydorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning & Technology, 14*(2), 50–73.
Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *Test of Word Reading Efficiency*. Austin, TX: Pro-Ed.
Van Heuven, W. J., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX–UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology, 67*, 1176–1190.
Wilkinson, K. S., & Houston-Price, C. (2013). Once upon a time, there was a pulchritudinous princess: The role of word definitions and multiple story contexts in children's learning of difficult vocabulary. *Applied Psycholinguistics, 34*, 591–613.
Williams, S. E., & Horst, J. S. (2014). Goodnight book: Sleep consolidation improves word learning via storybooks. *Frontiers in Psychology, 5*. https://doi.org/10.3389/fpsyg.2014.00184.