Evaluating the impact of vocabulary instruction on oral vocabulary, phonemic awareness and nonword reading

Rose Brooks[1] · Meesha Warmington[2] · Jenny Thomson[1]

© The Author(s) 2024

Abstract

Vocabulary teaching has traditionally focused on word meanings to aid reading comprehension, however evidence also suggests that vocabulary knowledge influences phonemic awareness and word reading. Vocabulary instruction concentrating on the sound structure of new words alongside meaning (combined sound-meaning vocabulary instruction) improves vocabulary for learners with speech and language difficulties. Emerging research suggests it may deliver similar results as a classroom strategy for young children. Researchers have questioned whether combined instruction would additionally enhance phonemic awareness and phonic decoding. A teaching intervention with 273 children aged 5–6 in the United Kingdom compared meaning-based instruction, combined instruction and usual classroom instruction (age-matched controls) over 24 weeks with three testing points. A daily vocabulary lesson incorporating evidence-based principles was delivered by classroom teachers. After intervention both instructional groups performed significantly higher than controls on taught vocabulary, and all groups performed equally on phonemic awareness and nonword reading. Delayed post-test results require more cautious interpretation due to the lack of a hierarchical design. Combined instruction resulted in the highest taught vocabulary and phonic reading outcomes; phonemic awareness in the combined condition was significantly higher than controls. Results tentatively suggest that combined instruction is an inclusive approach for whole-class use in early schooling.

Keywords Vocabulary · Phonemic awareness · Early literacy · Word reading

✉ Rose Brooks
r.brooks@sheffield.ac.uk

1   Department of Human Communication Sciences, University of Sheffield, Sheffield, UK

2   School of Education, University of Sheffield, Sheffield, UK

② Springer

Introduction

Vocabulary learning is vital to the development of oral language and literacy, so it is essential to investigate methods that will deliver maximal benefits. A combined method of vocabulary instruction which highlights the sound structure of words alongside the traditional emphasis on meaning has demonstrated efficacy for learners with low vocabulary levels, including those with speech, language and communication needs (German et al., 2012; Lowe et al., 2019) and those from economically disadvantaged backgrounds (Spencer et al., 2017; St John & Vance, 2014). Several initial studies have evaluated the approach for whole-class teaching, indicating greater vocabulary learning when instruction emphasises word form (Janssen et al., 2018; Silverman, 2007). If the combined sound-meaning approach can boost vocabulary learning in the classroom similarly or more than meaning-based (semantic) approaches, it could constitute an inclusive pedagogy benefitting a wide range of learners.

Vocabulary has traditionally been conceptualised and investigated in relation to reading comprehension (Elleman et al., 2009). However, a range of research suggests that oral vocabulary predicts phonemic awareness and word reading (e.g. Wagner et al., 1997). Further investigation is needed to evaluate whether vocabulary instruction has further impact on these early reading skills.

Impact of vocabulary instruction on oral vocabulary

Existing research indicates effective principles to optimise learning and retention of taught words. Explicit or direct vocabulary teaching supports a wide range of learners, producing nearly double the effect size of incidental vocabulary encounters (Marulis & Neuman, 2010). Younger students are less adept than older children at deducing word meanings from text (Biemiller & Boote, 2006). There is convincing evidence that interactive book reading and readalouds of high-quality storybooks offer an effective context for vocabulary teaching (Marulis & Neuman, 2010; Noble et al., 2020). Research favours the selection of vocabulary found in literary texts, pupil-friendly definitions, activity-based learning and multiple word encounters (Beck et al., 2013). Other strategies with positive research evidence include the use of visual images (Lawson-Adams & Dickinson, 2020) and a systematic cycle of vocabulary review (Bahrick & Hall, 2005).

Effect of semantic instruction on vocabulary

Although a number of meta-analyses have focussed on the impact of semantic vocabulary instruction on reading comprehension (e.g. Elleman et al., 2009), only one was discovered relating specifically to vocabulary outcomes. Marulis and Neuman (2010) examined 67 studies in the 5–6 year old age range and calculated a large effect size ($g = 0.88$) for taught vocabulary. Similar gains were made for whole class

($g = 1.04$), small group ($g = 0.88$) and individual ($g = 0.98$) delivery. An enduring issue for vocabulary interventions is the lack of generalisation to untaught words (ibid.).

Effect of combined instruction on vocabulary

An alternative approach focuses jointly on the phonological sound structure *and* semantic meaning of new words. Teaching of phonological form includes segmentation of words into larger and smaller units (e.g. syllable clapping, rhyme detection, alliteration). Semantic input often involves definitions, use in a sentence, examples and acting out the word. In addition to explicitly teaching the sound structure and meaning of words, combined instruction capitalises on the powerful associative link between these representations (Storkel & Morrisette, 2002).

The main difference between combined instruction and traditional semantic-based approaches relates to the explicit teaching of word structure. For example, robust vocabulary instruction (Beck et al., 2013) focuses on word meanings through a variety of contexts but not on word segmentation.

Emanating from the field of speech and language therapy, evidence suggests that combined instruction is effective for boosting targeted vocabulary for learners with Developmental Language Disorder (DLD) across the school age range (Ebbels et al., 2022; Lowe et al., 2019; Marulis & Neuman, 2010), and there is some indication that this also applies to cohorts from lower socioeconomic backgrounds (Spencer et al., 2017; St John & Vance, 2014). Whilst this body of research demonstrates enhanced outcomes of combined instruction over usual practice, investigation is still needed with a comparative semantic-only condition.

Several investigations have evaluated the combined approach for whole class vocabulary instruction, delivered by the class teacher using high quality storybooks. The majority of designs compare combined instruction to usual vocabulary practice only. A pre-test post-test design by Moran and Moir (2018) with 91 children aged 3–5 discovered significant improvement in vocabulary definitions ($p < 0.001$, no effect size reported). Quasi-experimental research by Damhuis et al. (2016) investigating 4–6 year olds in the Netherlands found significantly higher target vocabulary definitions in the combined condition compared to an age-matched control group with a large effect size ($N = 87$; $p < 0.001$; $\eta_p^2 = 0.28$). Using a similar design and sample Droop et al. (2005) observed significantly higher scores on a standardised expressive vocabulary measure for children in the combined intervention ($N = 223$; $p < 0.05$; $d = 0.23$) compared to age-matched controls. Several investigations have included a semantic-only comparative group. A study of 4–6 year olds ($N = 85$) in the Netherlands (Janssen et al., 2018) compared phonological and semantic instruction, however the lack of a control group and combined condition precluded evaluation of the combined strategy. Expressive definitions of taught words improved significantly more in the phonological group with a large effect size ($p = 0.001$, $\eta_p^2 = 0.14$). A robust classroom evaluation was carried out by Silverman (2007) with 94 children aged 5–6 across six US kindergarten classes in two schools. Testing occurred at three timepoints—pretest (T1), post-test (T2) and maintenance test (T3) six months later. Three experimental groups were included: (1) a combined sound-meaning

condition focussed on the phonemes and written letters of target words, (2) semantic instruction of word meanings and (3) discussion of target words in relation to children's own experience (contextual group), rather than a business-as-usual control group. At T2, the combined and semantic groups significantly surpassed the contextual group on target word definitions ($p < 0.01$) but were not significantly different to each other; the combined group had a higher effect size compared to both contextual instruction ($d = 1.19$) and semantic instruction ($d = 0.85$). At T3 with a reduced sample ($N = 50$) a significant difference was only detected on the definitions task, with the combined group outperforming the contextual group ($p = 0.01$; $d = 0.94$).

Impact of vocabulary instruction on word-level literacy

A number of researchers have called for studies to investigate whether combined vocabulary instruction leads to supplementary improvements in phonemic awareness and word reading beyond anticipated vocabulary gains (Dickinson et al., 2003; Duff et al., 2015; Munro et al., 2008). Several avenues of research point towards this possibility.

Many cross-sectional and longitudinal studies confirm that vocabulary size (amount) uniquely predicts phonemic awareness (Dickinson et al., 2003; Duff et al., 2015; Sénéchal, et al., 2006; Wagner et al., 1997) and word reading outcomes (Duff et al., 2015; Garlock et al., 2001; Lee, 2011; Wagner et al., 1997). Vocabulary size continues to be a stable predictor of phonemic awareness and word reading, particularly phonic decoding, until around age 8 as phonological skills typically reach maturity (Lee, 2011; Storch & Whitehurst, 2002; Wagner et al., 1997). Thereafter, vocabulary increasingly contributes to improvements in reading comprehension (Wagner et al., 1997) and exception word reading (Ouellette, 2006; Ricketts et al., 2007), for example the word 'yacht', which cannot be decoded without access to word meaning.

A plausible account of this relationship is offered by the lexical restructuring hypothesis (Metsala & Walley, 1998), proposing that phonological representations of words are initially stored as wholes, but as the lexicon grows in size, these become increasingly distinct and segmental to enable new words to be stored separately from existing items, thus forming the basis for explicit phonemic awareness needed for word reading. Empirical support for lexical restructuring derives from research confirming that phonological representations undergo a gradual process of refinement over the course of childhood (Ainsworth et al., 2015; Garlock, et al., 2001).

Also hinting towards the possibility of wider literacy outcomes is a body of literature challenging the traditional conceptualisation of reading as two distinct skills (language comprehension and word decoding), with decoding as a determinant of comprehension (Gough & Tunmer, 1986). Recent theory supports a more interrelated view of vocabulary and decoding as mutually supportive domains (Duke & Cartwright, 2021; Nation, 2019; Snowling & Hulme, 2020; Wegener et al., 2022).

Finally, several experimental studies of combined instruction which incorporated measures of phonemic awareness indicate that the combined approach may

additionally increase this skill. Theoretical justification is found in the lexical quality hypothesis (Perfetti & Hart, 2002), predicting that word learning is enhanced by attention to multimodal features such as sound, meaning and print. Analysing the sound structure of new vocabulary could provide the underpinnings for explicit phonemic awareness (Metsala & Walley, 1998), and direct teaching of phonemic awareness is a known causal factor in word reading (Ehri et al., 2001).

In a pretest post-test design, Munro et al. (2008) administered a combined vocabulary programme individually to 17 Australian preschoolers with DLD (ages 4–6) in a speech and language clinic. Significant post-intervention gains ($p < 0.05$) with large effect size were found for rhyme recognition ($\varepsilon^2 = 0.66$) and alliteration ($\varepsilon^2 = 0.63$). The design was extended by Coloma et al. (2022) with a larger sample ($N = 43$) of 5–6 year old preschoolers in Chile and a control group. Syllable awareness was chosen as a developmentally appropriate measure of phonological awareness due to the age of the sample. Significantly higher performance was discovered at post-test in the combined intervention group compared to controls, $t(41) = 2.81$, $p = 0.008$. Janssen et al. (2018) compared separate phonological and semantic conditions and consequently discovered significantly higher post-test scores in the phonological group on an early literacy measure including rhyme recognition and phoneme blending with a medium effect size ($p = 0.02$, $\eta_p^2 = 0.08$).

No peer-reviewed papers have yet measured the effect of combined vocabulary instruction on word reading accuracy, although connectionist theories of reading (e.g. Harm & Seidenberg, 2004) provide some theoretical support for this prospect. Other empirical research highlights a role for both phonological and semantic facilitation in word reading (Ouellette & Fraser, 2009).

The present study

Previous research indicates that combined vocabulary instruction is an effective approach to boost vocabulary for specific cohorts with low levels of vocabulary compared to usual instruction. A small number of studies suggest that combined instruction can also be used as a whole-class strategy, showing improved target vocabulary compared to either a semantic or a control group.

To further our understanding of whether combined instruction is a viable classroom approach, it will be useful to build upon the design by Silverman (2007) which included three groups and three timepoints, but this time incorporating a business-as-usual control group for comparison to the combined and semantic conditions as well as a larger T3 sample. Whilst initial studies indicate the possibility of phonemic awareness gains arising from vocabulary instruction (Janssen et al., 2018; Munro et al., 2008), this is an early line of research requiring considerable further evaluation in the classroom with stronger research designs.

The current quasi-experimental study therefore aims to extend the literature by investigating the impact of whole-class vocabulary instruction with 5–6 year olds on oral vocabulary, whilst incorporating measures of phonemic awareness and phonic decoding to begin to consider whether these might be additionally affected by vocabulary instruction. The early school years are an optimal time to capitalise

on the relationships between vocabulary, phonemic awareness and word reading that exist in this age group (Wagner et al., 1997). Whole-class instruction is an effective option, yielding outcomes equivalent to small-group interventions (Marulis & Neuman, 2010). It is also an expedient choice, given the large cohorts of children with limited vocabulary (Speech & Language UK, 2023).

Accordingly, the research questions ask which of three vocabulary teaching approaches (combined, semantic, control) results in the highest performance in vocabulary, phonemic awareness and phonic reading in 5–6 year olds.

Hypothesis 1 At T2 both teaching groups are expected to perform equally (and better than controls) on taught vocabulary due to equivalent dosage of teaching input. No significant improvement is expected from T2 to T3 in any group since these items were not taught during this period.

Hypothesis 2 No significant group differences are predicted on standardised vocabulary at any time point, in accordance with other intervention studies failing to demonstrate distal gains beyond taught items (Marulis & Neuman, 2010).

Hypothesis 3 At T2 and T3 the combined group is expected to perform significantly higher than both other groups on phonemic awareness owing to the explicit teaching of phonological segmentation.

Hypothesis 4 Nonword reading outcomes are expected to mirror results for phonemic awareness at both T2 and T3, given the anticipated mediating effect of phonemic awareness on phonic reading.

Method

Research design

The study was approved by the ethics review panel of the University of Sheffield, Department of Human Communication Sciences. The investigation was conducted during the 2018–19 academic year with 273 children aged 5–6 based on an a priori power analysis. In the UK, children enter school and begin literacy instruction at age 4–5, so the current sample is in the second year of schooling (Year One).

The current quasi-experimental design sought to answer the research questions by incorporating three teaching conditions (semantic, combined and usual practice) and three timepoints for testing: T1 in September directly before teaching, T2 in June/ July post intervention and T3 four months later in November of the next academic year. The waiting control group received the programme after the T3 data was collected. The age-matched control group was incorporated to demonstrate the effect of maturation and usual (mainly incidental) vocabulary teaching. The group receiving meaning-based training illuminated the impact of traditional high-quality semantic

teaching, whereas the combined group indicated the supplementary effect of explicit teaching of phonological form.

Classes took part in a daily vocabulary teaching linked to storybooks over the course of 24 weeks, after a two-week trial period. In September prior to the programme, a teacher questionnaire was administered to collect data on teacher characteristics, and a half-day training session was provided for intervention teachers.

The sample

Recruitment

Schools were invited to volunteer for the research programme if they met the following eligibility criteria: (1) school location within an hour's travel for the researcher and testers and (2) average class size of at least 25 in an effort to recruit single-age classes of 5–6 year olds (rather than mixed-age classes) to support implementation of the whole-class teaching programme.

School characteristics

Sixteen classes across nine schools participated in the study, located within urban, suburban and rural settings spanning a wide socio-economic spectrum (see Table 1). The first nine enrolled classes entered the intervention arm and were randomly allocated to the semantic and combined instructional groups. Classes from each school remained in the same instructional group to avoid exposure to the alternate teaching approach. Later recruits became the waiting control group, which included two large schools with seven mixed-age classes (no further single-age classes came forward), however only the 5–6 year old classes were assessed.

Table 1 School characteristics

	Setting	Number of classes	Number on roll	IMD decile	FSM %	Group allocation
School A	Suburban	two	412	8	7.5	Combined
School B	Rural	one	198	10	10.1	Semantic
School C	Suburban	two	420	9	9.9	Semantic
School D	Rural	one	189	4	23.3	Semantic
School E	Urban	one	235	3	43.4	Combined
School F	Urban	one	211	2	47.8	Semantic
School G	Urban	one	377	4	37.6	Combined
School H	Rural	five*	438	10	19.9	Control
School I	Urban	two**	335	2	38.6	Control

Number on roll excluding nursery (GOV.UK, 2018)

FSM Free school meals (ibid.), *IMD* Index of multiple deprivation (GOV.UK, 2015)

*Five mixed-age classes equivalent to two classes

**One single-age and one mixed-age class

Given that full randomisation was not achieved, additional analyses were performed on influential school and teacher-level variables to ascertain whether the later-recruited control group differed from the taught groups in ways that could affect results. Table 1 establishes that each group covered a broad range on the socioeconomic indices. Responses to the teacher questionnaire indicated equivalent teacher motivation and confidence as seen in Table 2 and Table 3.

The Results section will explore further potential differences between participant groups arising from group allocation.

Learner characteristics

A sample of 124 girls and 149 boys participated in the study with a mean age of 5 years 6 months (*range* = 5;1–6;0, *SD* = 3.31). In total, 89% of parent/carer consent forms were returned, resulting in 278 potential recruits. Five exclusions were made at T1 due to significant special needs, yielding the required sample size. Attrition from T1 to T3 amounted to nine participants (*N* = 264). The sample comprised 96% monolingual speakers, and 21% had identified special educational needs. 13.2% of learners came from economically disadvantaged backgrounds, based on the pupil premium funding received by schools for children on low income.

Teacher characteristics

Nine teachers (six female and three male) with 1–14 years of teaching experience (*M* = 4.9) took part in the teaching intervention, and a further seven teachers (six female and one male) joined the control group with 2–10 years of experience (*M* = 5.2). According to the teacher questionnaire, usual practice across all groups consisted predominantly of incidental vocabulary discussion, a pattern confirmed in research by Blachowicz et al. (2006). Additional strategies used by individual teachers can be seen in Table 4.

Table 2 How important is vocabulary teaching? (Out of 5 points: 1 = Not important, 5 = Very important)

	N	M	SD	Min	Max
Taught classes	9	4.56	0.527	4	5
Control classes	7	4.50	0.707	44	55

Table 3 How confident do you feel about teaching new vocabulary? (Out of 5 points: 1 = Not confident, 5 = Very confident)

	N	M	SD	Min	Max
Taught classes	9	3.00	0.866	2	4
Control classes	7	3.00	0.000	3	3

Table 4 Vocabulary practice prior to intervention

	Incidental encounters	Images	Word bank/wall	Preteaching subject words
Semantic group	3	1	2	0
Combined group	3	0	0	0
Control group	5	0	0	2

In terms of prior training, five schools had received a staff meeting on vocabulary (two semantic, two combined, one control), while six schools had received no training (three semantic, two combined, one control), indicating an even balance.

Assessment measures

The assessment battery was trialled and modified during an earlier pilot study.

Vocabulary

Two standardised vocabulary assessments were included to capture the size of the child's receptive and expressive vocabulary. In the British Picture Vocabulary Scales 3 (BPVS3; Dunn et al., 2009) a plate of four coloured pictures of increasing difficulty is displayed. The assessor says a word and asks the child to point to the corresponding picture (test reliability $\alpha = 0.91$). The Clinical Evaluation of Language Fundamentals 4 Expressive Vocabulary subtest (CELF4; Semel et al., 2003) depicts 27 objects and actions of increasing age of acquisition. Children continue naming the images until seven consecutive errors are made (test reliability $\alpha = 0.84$).

Consistent with most vocabulary studies (Marulis & Neuman, 2010), a definitions task was devised by the researcher to assess taught vocabulary, capturing both vocabulary size and depth. A set of 21 randomly selected target words was extracted from the full list of 108 items (19%), approximately two from each storybook (see Appendix A). Independent samples t tests demonstrated no significant differences on age of acquisition (Kuperman et al., 2012), written word frequency (Kucera & Francis, 1967), word length (number of phonemes) or phonotactic probability, i.e. the relative frequency of sound segments in a word (Vitevitch et al., 1999), thus indicating that the assessment was a valid representation of the full vocabulary set. Both sets were predominantly composed of nouns, verbs and adjectives.

The tester asked the child to provide the meaning of each word and wrote this verbatim onto the record sheet for later scoring. An earlier trial confirmed that audio-recording was not necessary since tester transcription was fully accurate. A scoring matrix was created using an iterative process until all responses were acknowledged. The researcher and a Speech and Language Therapist/Pathologist each marked the definitions test according to the matrix. Two points were awarded if a clear understanding was demonstrated, 1 for a partial or imprecise response and 0 when no understanding was shown. Afterwards, an inter-rater reliability (IRR) check was

performed on 15% of the sample at T1 (41 students) to measure consistency between the two markers. The IRR of 92% agreement fell in the substantial range (Cohen's kappa $=0.76$). Discrepant items were moderated between the two markers, resulting in agreement for all items.

Phonemic awareness

This was measured through the standardised Comprehensive Test of Phonological Processing 2 Elision subtest (CTOPP2; Wagner et al., 2013) which taps the ability to delete a phonological segment from a spoken word to create a new word (test reliability $\alpha = 0.92$). The first nine items tested syllable deletion (*say sunshine without sun*), however the majority of the test focussed on phoneme deletion (*say farm without saying /f/*), leading to the decision to consider this as a measure of phoneme deletion. Elision was chosen due to its relatively high age norms (Yopp, 1988) to minimise ceiling performance effects. It is recognised that the task presents a high working memory demand since the target is held in memory whilst deleting the initial phoneme. Alliteration and rhyme were tested using the Phonological Awareness Battery 2 (PhAB2; Gibbs & Bodman, 2014), although these will not be reported due to sizeable ceiling effects.

Phonic reading

The PhAB2 Nonword Reading Test (Gibbs & Bodman, 2014) requires the child to read nonwords of increasing length (test reliability $\alpha = 0.84$). Nonwords minimise reliance on stored information to support word reading (Roodenrys & Hinton, 2002), thus representing an optimal measure of phonic decoding. Phonic decoding was chosen to measure reading accuracy due to its prominence in the relationship between vocabulary, phonemic awareness and word reading (Wagner et al., 1997). IRR was checked with 15% of the T1 sample, since careful discrimination of children's oral responses is required, e.g. */rad/* versus */red/*. Five testers each listened to audio recordings for eight randomly selected students from a different tester, totalling 40 students. IRR fell in the substantial range (Cohen's kappa $=0.76$) representing 89% agreement.

Assessment procedure

At each testing point, participants completed two individual 20-minute sessions, administered on separate days within the same week. Each testing point lasted around three weeks. Assessment took place in a quiet area of the school with testers unaware of group allocation. All were skilled in working with young children, i.e. qualified teachers, Speech and Language Therapists/Pathologists and psychology Masters students. Testers received two individual two-hour training sessions, culminating in a test and follow-up practice until 100% accuracy was achieved. The researcher observed testers on the first day of testing at each time point to monitor

adherence to the protocol. The vocabulary intervention was implemented from the end of T1 testing until the start of T2 testing.

Teaching materials

Evidence-based principles drawn from the research literature provided a strong basis for instruction in both taught groups.

Materials and procedures were trialled for two weeks during a previous pilot study. All teachers received identical sets of intervention resources, differing only in the type of vocabulary facilitation cue (semantic or combined) on the teaching card and games. The following manualised scheme of intervention materials was provided (in print and electronic format) to aid consistency and fidelity. *Teaching protocol.* A concise two-sided sheet explained the standard teaching protocol. *Storybooks.* Twelve engaging storybooks and a trial book (in Appendix B) were selected from fiction lists for 5–6 year olds (e.g. CLPE, 2018) and sourced for each class. *Teacher planning records.* A simple plan was provided for each two-week unit with the target words, pupil-friendly definitions and review vocabulary. Definitions were created using the Wordsmyth (2014) website and included common tier one vocabulary according to principles set out by Beck et al. (2013). Teachers were asked to note completed lessons each day and to provide optional feedback at the end of each week. *Symbolised vocabulary cards.* Nine words were selected by the researcher from each book to span a two-week teaching unit (total of 108 words) largely based on the tier two categorisation of Beck et al. (2013). During the pilot study 87% agreement was reached on tier two word selection between the researcher and class teachers. The programme did not teach tier one high frequency words which form part of everyday oral language nor tier three low frequency words linked to a specific subject or topic (ibid.). Age of acquisition norms from 5 to 10 years (Kuperman et al., 2012) were included to ensure suitability for a wide ability range and to mitigate ceiling effects. Exclusions were made for multiword phrases, words unfamiliar to the UK context and low frequency (tier three) words. Symbolised vocabulary cards were created in colour using Widgitonline (Widgit, 2007) with text underneath on a grid of nine per page (see Appendix C). *Teaching cue card.* Separate cards (available in Appendix D) were created for the semantic and combined groups using Widgitonline (Widgit, 2007) as a large colour poster and in digital format. *Menu of games.* Five simple practice games (described in Appendix E) were created by the researcher based upon activities by Parsons and Branagan (2014). A set of 10 fully prepared games (two of each type) was given to each class. *Word wall resources.* Each class set up a display to facilitate application of the taught vocabulary comprising symbolised word cards for the unit, a template to showcase the word of the day, a pocket chart for storing the taught vocabulary cards, a voice level thermometer to support quiet talking during the game and space to display student work.

Teaching procedure

All Groups

Literacy instruction for all groups followed the UK National Curriculum, involving daily literacy activities linked to a class text (about one hour). In the first three years of schooling, there is also a separate short daily phonics lesson, culminating in a phonics test at the end of Year One (age 5–6; NFER, 2013).

Usual teaching curriculum (control group)

According to the teacher questionnaire, vocabulary in the control group was discussed incidentally during the readaloud. Two teachers responded that they additionally carried out preteaching of subject vocabulary (tier three words in the hierarchy of Beck et al., 2013), which was not the case for intervention teachers.

Groups receiving the vocabulary teaching programme (combined and semantic groups)

The programme contained 12 two-week teaching units each based on a storybook (24 weeks, 20 hours of instruction), plus an untested two-week trial. In common with other vocabulary intervention studies, students were taught one vocabulary item per day (Marulis & Neuman, 2010).

At the start of each unit, the teacher read the specified storybook aloud to the class. Most classes used the programme storybooks as their class text, although two schools (one combined, one semantic) had already planned their story texts for the year and therefore chose to implement the programme at a different time of day. After the initial readaloud, teachers revisited the book each day as part of their literacy instruction throughout the two-week block, highlighting taught words as they arose in context. To minimise variation between classes, teachers were requested not to explicitly teach any other vocabulary during the intervention period nor to send home vocabulary practice activities. In case of teacher absence or special events, teachers could either facilitate their teaching assistant to deliver the lesson or double up on the target words the following session.

The daily vocabulary lesson lasted approximately 10 minutes: four minutes teaching input, four minutes for a game and a two-minute review. Children sat on the carpet facing the teacher and the vocabulary teaching card (semantic or combined). Vocabulary was taught using the four-step STAR protocol (Blachowitz et al., 2006). *Select*. Tier two word selection was described in the intervention materials. *Teach*. Day one of each new unit familiarised children with all nine words. Children were asked what they knew about the words, followed by the teacher reading a simple pupil-friendly definition from the planning sheet provided. On the other nine days, the teacher taught the word of the day by pointing to each of the six cues on the teaching card (in

Appendix D), followed by question–answer or paired pupil discussion. The semantic teaching card contained three facilitation cues (meaning, sentence, acting out) each used twice. To evaluate the additional effect of phonological form on the research outcomes (over the same time frame), the combined teaching card included the three semantic cues plus three phonological cues (rhyme/syllables, phoneme counting, clear articulation). To support a wide range of learners, suggested wording was provided that included at least six exposures to the word (as suggested by McGregor et al., 2021). By way of an example, to teach the word 'cluster' in the combined group the teacher would point to each learning cue in turn and say: *What does 'cluster' mean? Can you use 'cluster' in a sentence? Let's say 'cluster' clearly. Who would like to act out the word 'cluster'? Let's clap out the syllables in 'cluster' (or tell me a word that rhymes). Let's sound out the phonemes in 'cluster' on our fingers.* Teachers were also encouraged to highlight morphemic variations during the teaching session, e.g. crawls/crawled/crawling, as this has been shown to enhance both oral vocabulary and literacy (Breadmore et al., 2021). They were requested not to draw attention to the printed word to minimise the additional orthographic variable. *Apply.* After the teaching input, children went to their tables to play a game for four minutes in small groups of 2–4 to practise the word of the day. Teachers instructed students how to play the games, supporting small-group interaction as needed. Measures were taken to minimise background noise to enable spoken vocabulary to be heard clearly, including foam dice and a voice thermometer. *Review.* Children returned to the carpet for a two minute review of the word of the day, the previous day's word and the word from a week prior, premised on research showing that distributed practice enhances word retention more than consecutive practice (Carpenter & DeLosh, 2005) and that an expanding retrieval schedule leads to higher retention than fixed intervals (Bahrick & Hall, 2005). Children gave a definition or sentence for the review words to optimise expressive vocabulary use. The word wall was available throughout the day to boost application of the taught vocabulary.

Teacher training session

Intervention teachers received a two-hour individual training session (see Table 5) during the initial testing period. If schools had two classes, the session was delivered jointly to both teachers. There was an opportunity to ask questions, and teachers

Table 5 Content of teacher training session

1	Teacher questionnaire
2	Project overview (alternative approach not discussed)
3	Familiarisation with the story books and target vocabulary
4	Classroom organisation and setting up the word wall
5	Demonstration of the STAR protocol
6	Exploration of programme resources
7	Time to play each of the games
8	Discussion of classroom management, consistency checks and diary dates

could contact the researcher at any time during the programme. Participating teachers were unaware that two approaches were being compared, which does not present an ethical dilemma since both methods have evidence of efficacy (Silverman, 2007).

Treatment fidelity

Researcher visits were carried out four times during the programme. The first was an informal observation during the trial unit to offer guidance. The other three (spaced out over the year) were scored to monitor adherence to the teaching protocol. Ten intervention components were observed and scored as 0 or 1 (see Table 6). Development points and positive observations were discussed with the teacher. A score of 9 or 10 was considered excellent, a score of 8 was good, and scores of 7 and below prompted a more in-depth discussion and review of the protocol. Discussion points centred around the need to adhere to the specified timings and reducing noise levels during the game. Teachers' viewpoints and suggestions were also gathered during these visits.

A high mean consistency rating across observations ($M = 9.5$; $SD = 0.333$; *range* 7–10) indicated that the intervention was delivered according to the protocol.

Teacher questionnaire

A six-item teacher questionnaire was administered prior to the programme to gather information from participating teachers. The first two questions used a Likert scale from 1 to 5 (1 = low, 5 = high). The rest were free-field questions allowing open responses (see Table 7).

Data previously presented in Tables 2–4 compared teacher characteristics that could potentially influence study results.

Table 6 Fidelity checklist

1	Target vocabulary taught daily; missed lessons noted on planning sheet
2	Teaching card and word of the day displayed at start of lesson
3	Words taught using the six cues on the teaching card
4	Approximate timings (in minutes) followed: 4 teaching, 4 game, 2-review
5	Pupil-friendly definition provided
6	Games checked fortnightly and replenished as needed
7	Noise levels monitored during game time
8	Pupils play each of the games during the two-week unit
9	Words reviewed according to the planning sheet
10	Word wall is updated daily with target word and fortnightly with new vocabulary set

Table 7 Teacher questionnaire items

1	How important is vocabulary teaching? (1–5)
2	How confident do you feel about teaching new vocabulary? (1–5)
3	Can you describe the vocabulary teaching and activities you have used?
4	What are the main barriers to classroom vocabulary instruction in your view?
5	How many years have you been teaching, and in which year groups?
6	What training opportunities have you received to support the teaching of vocabulary?

Results

Data analysis plan

A oneway ANOVA indicated a significant group difference on the IMD (socioeconomic indicator), $F(2,260)=3.981$, $p=0.02$, $\eta_p^2=0.03$; therefore group equivalence was examined for all measures (see Table 8).

A statistically significant difference was found only on the BPVS3 (receptive vocabulary) raw scores, $F(2,270)=4.254$, $p=0.02$; $\eta_p^2=0.03$. Post hoc analysis with Bonferroni correction indicated that the semantic condition was significantly higher than the combined condition ($p=0.03$ for IMD; $p=0.01$ for BPVS3). To reduce variance and improve reliability, the pretest BPVS3 score was entered as a covariate in ANCOVA, in the tradition of other vocabulary research (Damhuis et al., 2016; Janssen et al., 2018).

Repeated measures ANCOVA with pretest BPVS3 as covariate was performed on raw scores for taught vocabulary, phonemic awareness and nonword reading with instructional *group* as the between-subjects factor (semantic, combined, control) and *time* as the within-subjects factor (T1, T2, T3).

Prior to analysis, procedures were followed for data screening, normality checks and assumptions. The Levene's statistic was consulted for homogeneity of variance at all timepoints, however since this statistic is often inflated in large samples (Field, 2013), and since ANCOVA with similar group sizes is robust to this violation (ibid.) significant results were followed with a calculation of the variance ratio, dividing the largest group variance by the smallest to check that the result was less than three (Jaccard, 1998). Linearity was confirmed through visual inspection of scatterplots between the T1 BPVS3 covariate and the outcome variable at each timepoint. Homogeneity of regression slopes (HRS) was determined by scrutiny of

Table 8 Group differences on outcome measures

Variable	F	df 1	df 2	Sig
CTOPP2 Elision	1.218	2	270	0.30
PhAB2 Nonword Reading	2.185	2	267	0.12
BPVS3 Receptive Vocabulary	4.254	2	270	0.02*
CELF4 Expressive Vocabulary	2.772	2	270	0.06
Taught Vocabulary Definitions	2.281	2	268	.10

*Significant at $p=.05$

the interaction term. If a significant interaction was detected, indicating an uneven influence of the covariate across groups, separate ANCOVAs were run for the time-point in question to assess homogeneity in each pair. HRS was assumed if regression slopes were equal for two of the pairs (ibid.).

Vocabulary outcomes

Means (M), means adjusted for the T1 BPVS3 covariate (M_{adj}), standard deviations (SD), minimum scores (Min) and maximum scores (Max) for taught vocabulary definitions are presented in Table 9.

ANCOVA controlling for pretest BPVS3 resulted in a statistically significant time x group interaction, $F(4,516)=60.032$, $p<0.001$, $\eta p^2=0.32$, large effect size. Post hoc pairwise comparisons with a Bonferroni correction found no significant group differences in taught vocabulary at T1 (combined-semantic $p=1.00$; combined-control $p=0.09$; semantic-control $p=0.40$). At T2 the combined and semantic groups displayed comparable results ($p=1.00$), with both the combined group ($p<0.001$, $d=1.06$, large effect) and semantic group ($p<0.001$, $d=1.01$, large effect) achieving significantly higher than controls. At T3, the combined teaching group knew significantly more taught vocabulary than both the control group ($p<0.001$, $d=1.78$, very large effect) and the semantic group ($p<0.001$, $d=0.54$, medium effect). The semantic group also performed significantly better than controls ($p<0.001$, $d=1.15$, large effect). Results are depicted in Fig. 1.

No treatment effects were detected on standardised vocabulary assessments. A mixed ANOVA on the BPVS3 data resulted in a significant time x group interaction, $F(3.858,501.553)=3.850$, $p=0.005$, $\eta_p^2=0.03$, small effect size, however post hoc pairwise comparisons with a Bonferroni correction discovered only the significant pretest difference discussed previously. ANCOVA results for CELF4 Expressive Vocabulary indicated no significant time x group interaction, $F(4,518)=2.034$, $p=0.09$; $\eta_p^2=0.02$ so post hoc analyses were not run.

Table 9 Taught vocabulary definitions by group (out of 42 points)

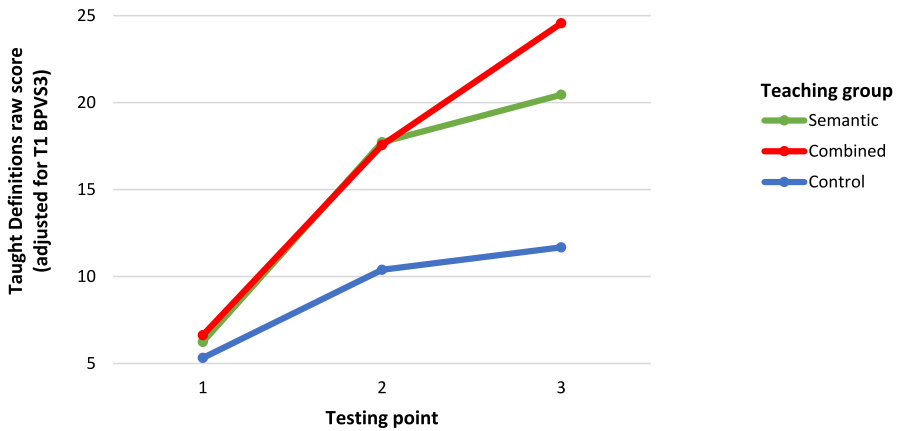|		N	M	M_{adj}	SD	Min	Max
Semantic group	T1	89	6.88	6.24	4.819	0	26
	T2	90	18.90	17.72	8.276	0	33
	T3	88	21.75	20.45	7.947	2	36
Combined group	T1	95	6.18	6.63	3.989	0	18
	T2	95	16.52	17.56	7.277	0	31
	T3	94	23.53	24.56	7.243	4	37
Waiting controls	T1	80	5.21	5.32	4.741	0	18
	T2	81	10.22	10.39	6.099	0	25
	T3	81	11.52	11.68	7.267	0	30

Fig. 1 Taught definitions ANCOVA outcomes by group

Phonemic awareness outcomes

Group outcomes for CTOPP2 Elision are presented in Table 10.

ANCOVA controlling for pretest BPVS3 detected a significant time x group interaction, $F(3.695, 480.337) = 6.951$, $p < 0.001$, $\eta_p^2 = 0.05$, medium effect size. Post hoc pairwise comparisons with a Bonferroni correction found no significant group differences on elision at T1 (combined-semantic $p = 1.00$; combined-control $p = 0.94$; semantic-control $p = 0.55$) or at T2 (combined-semantic $p = 0.67$; combined-control $p = 1.00$; semantic-control $p = 0.44$). At T3 there was a significant difference between the combined group and controls (($p = 0.002$, $d = 0.5$, medium effect size) but not between other groups (combined-semantic $p = 0.12$; semantic-control $p = 0.51$). Results are shown in Fig. 2.

Table 10 CTOPP2 Elision outcomes (out of 34 points) by group

		N	M	M_{adj}	SD	Min	Max
Semantic group	T1	90	8.94	8.10	6.150	0	32
	T2	90	15.31	14.57	7.135	0	33
	T3	89	17.45	16.69	7.750	0	33
Combined group	T1	95	7.92	8.57	5.748	0	28
	T2	95	15.16	15.83	7.197	0	29
	T3	94	18.22	18.84	6.386	0	33
Waiting controls	T1	81	9.26	9.37	6.121	0	24
	T2	81	15.99	16.10	7.955	0	32
	T3	81	15.09	15.20	8.543	0	32
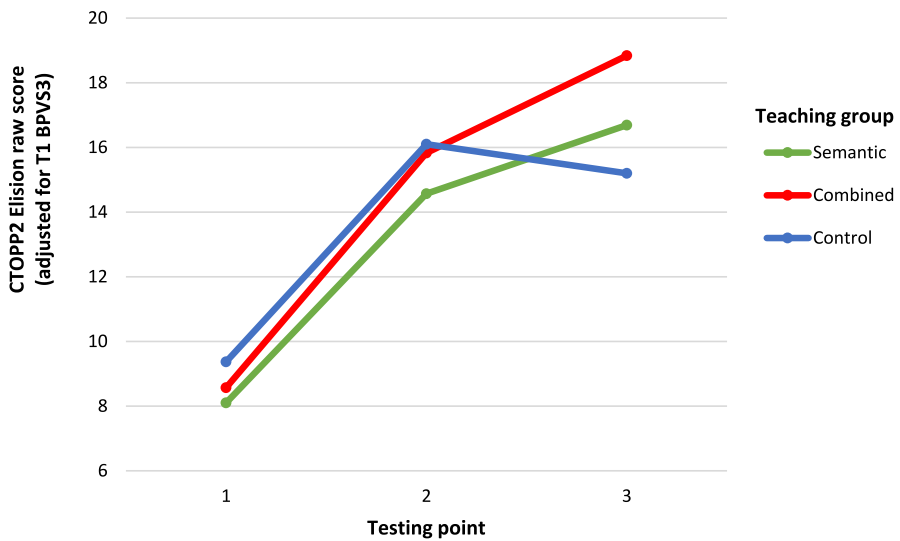
Fig. 2 CTOPP2 Elision ANCOVA outcomes by group

Nonword reading outcomes

Group outcomes for the PhAB2 Nonword Reading are presented in Table 11.

ANCOVA controlling for pretest BPVS3 found a statistically significant time x group interaction, $F(3.762, 483.387) = 11.593$, $p < 0.001$, $\eta_p^2 = 0.08$, medium effect size.

Post hoc pairwise comparisons with a Bonferroni correction indicated no significant group differences in nonword reading at T1 (combined-semantic $p = 1.00$; combined-control $p = 0.40$; semantic-control $p = 1.00$) or at T2 (combined-semantic $p = 0.77$; combined-control $p = 1.00$; semantic-control $p = 1.00$). At T3, the combined group performed significantly better than the semantic group ($p < 0.001$, $d = 0.67$, medium effect size) and controls ($p = 0.005$, $d = 0.49$, medium effect size). There was no significant difference between the semantic and control groups ($p = 0.81$). Results are displayed in Fig. 3.

Table 11 PhAB2 Nonword reading outcomes (out of 24 points) by group

		N	M	M_{adj}	SD	Min	Max
Semantic group	T1	87	6.23	5.76	4.574	0	19
	T2	90	13.12	12.53	5.382	0	24
	T3	89	13.47	13.00	6.533	0	24
Combined group	T1	95	5.13	5.51	4.273	0	16
	T2	95	13.06	13.44	4.957	0	24
	T3	94	16.46	16.81	4.830	1	24
Waiting controls	T1	81	6.52	6.57	5.109	0	23
	T2	81	12.81	12.87	6.333	0	24
	T3	81	13.94	14.00	6.731	1	24
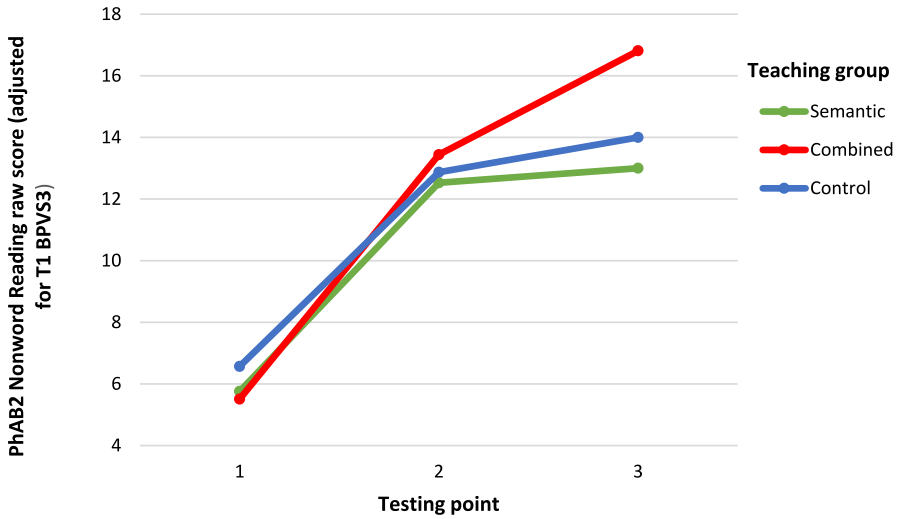
Fig. 3 PhAB2 Nonword reading ANCOVA outcomes by group

Discussion

The current investigation set out to determine the impact of vocabulary instruction with and without attention to phonological form on three outcomes of high educational importance, i.e. vocabulary, phonemic awareness and phonic reading.

As hypothesised, the strong instructional design and equivalent dosage supported equal results for taught vocabulary in the semantic and combined groups directly after intervention at T2, coinciding with classroom findings by Silverman (2007). This suggests equivalent vocabulary outcomes regardless of teaching method. We can have the most confidence in the T2 outcomes as an immediate test of the intervention condition. T3 outcomes should be interpreted with caution due to the lack of clustering, which could mean that school/teacher differences may have influenced the differential growth between T2 and T3. At T3 superior definitions performance (albeit with large variation) was found in the combined group compared to both other groups, and the semantic group also learned significantly more vocabulary than controls. A future study with a hierarchical design is needed to determine whether the difference at T3 was an effect of teacher/school factors or whether it may reflect the theoretical impact of higher lexical quality (Perfetti & Hart, 2002) and more segmental phonological representations (Metsala & Walley, 1998).

Group differences were not seen on the standardised vocabulary assessments, in line with our hypothesis and mirroring the endemic problem in vocabulary research of global measures being less sensitive to vocabulary gains (Marulis & Neuman,

2010). Vocabulary generalisation remains an important and continuing goal for future research. In the meantime, the accumulation of tier two vocabulary through direct word-a-day teaching linked to a storybook context could amount to considerable increases over the course of a child's schooling.

The hypothesised phonemic awareness and nonword reading advantage was not confirmed at T2, when all groups performed equally, ostensibly linked to the strong effect of classroom phonics teaching leading up to the end of year testing. The following academic year (T3), the combined group scored significantly higher on phonemic awareness compared to controls and significantly higher than both groups on nonword reading. Whether this is related to the mediating effect of phonemic awareness input (Ehri et al., 2001) in the combined condition or a result of school/teacher factors is an avenue for future enquiry.

The main limitation of the present study lies in the need for a hierarchical nested design to account for variance in outcomes linked to school and class-level variables, particularly the important effect of teaching style. This would necessitate an upscaled sample of schools. A further potential constraint relates to the largely monolingual sample. Generalisability to other school populations with more diverse and multilingual cohorts should be carefully considered, and indeed wider populations should be included in future research. A third limitation is the lack of full randomisation (early applicants joined the teaching groups but not the waiting control group) which could introduce participant bias if initial recruits had higher motivation levels. Whilst this possibility cannot be discounted, scrutiny of the teacher questionnaire indicates that teachers attached similar importance to vocabulary teaching and had similar levels of training. A fourth limitation is the level of word difficulty (AoA) for some selected words, which was included to avoid ceiling performance for data analysis. Best practice would suggest word choice closer to the children's experience to aid application.

Conclusion

The current study provides preliminary evidence that an integrated approach to teaching vocabulary with a dual emphasis on sound structure and meaning enhances vocabulary growth in the mainstream classroom for younger learners. Phonemic awareness and phonic reading may additionally be affected, although further testing is needed with a hierarchical design. Based on existing theory and research, reading instruction should acknowledge the reciprocal nature of decoding and comprehension skills (Duke & Cartwright, 2021; Nation, 2019). Current results favour consideration of combined vocabulary instruction as an inclusive classroom approach to stimulate growth in vocabulary and potentially to supplement early reading skills pending the outcome of further studies.

Appendix

Appendix A *Randomly selected vocabulary for definitions test*

searched
realised
spy
muscles
distracting
wobble
dunes
noticed
ancient
choir
squawk
grab
lair
tumbling
stroll
personality
disguise
dive
mysterious
tool
anchor

Appendix B *Storybooks*

Title	Author
Naughty Bus	Trial book kindly donated by Jan and Jerry Oke
Augustus and his SmileAugustus and his Smile	Catherine Rayner
How to Babysit a Grandad	Jean Reagan
How to Catch Santa	Jean Reagan
Don't Spill the Milk	Stephen Davies and Christopher Corr
How to Hide a Lion at School	Helen Stephens
Could a Penguin Ride a Bike?	Camilla de la Bedoyere
The Day Louis Got Eaten	John Fardell
Previously	Alan Ahlberg
Wanted the Perfect Pet	Fiona Robertson

Title	Author
Traction Man is Here	Mini Grey
Mrs Armitage on Wheels	Quentin Blake
The Sand Horse	Ann Turnbull

Appendix C *Sample of symbolised vocabulary cards for a two-week unit*
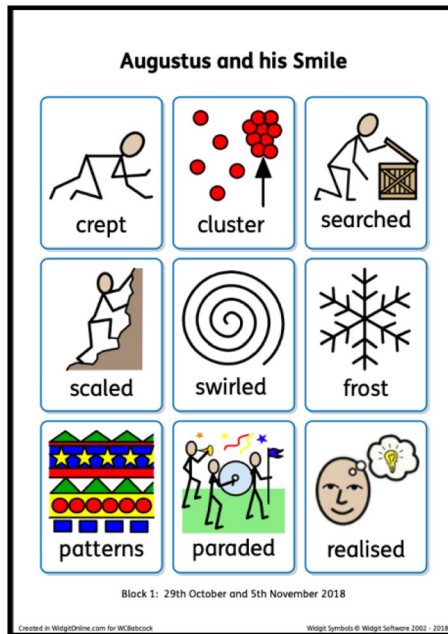
Widgit Symbols ©2002-2022

Appendix D *Teaching cue cards*
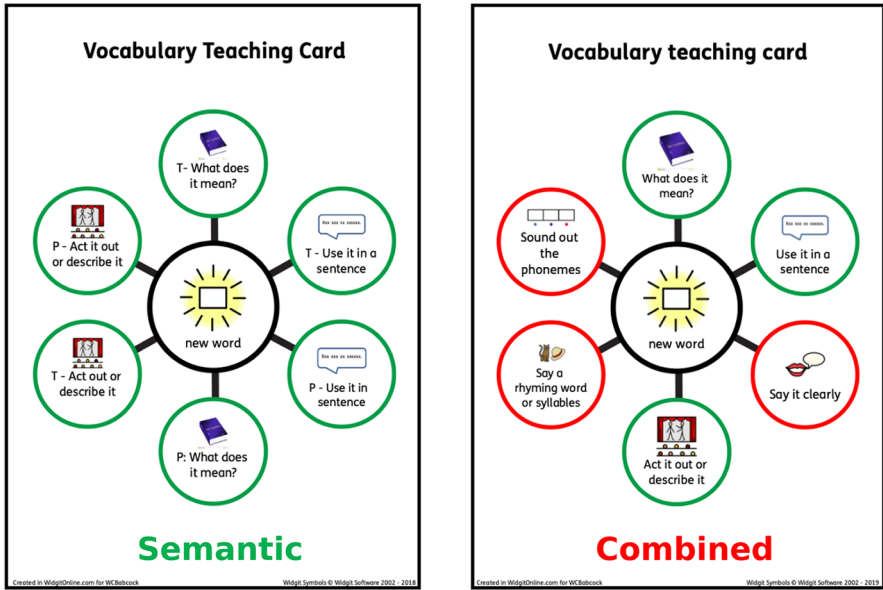
Adapted with permission from Parsons and Branaghan (2014) using Widgit Symbols ©2002–2022.

Appendix E *Game descriptions* (linked to learning cues in Appendix D)

Beetle game. Child rolls the dice, completes the learning cue and draws the indicated part of the beetle until complete.

Dice game. Child carries out the cue indicated by the dice.

Spinner game. Child carries out the cue indicated by the spinner.

Vocabulary swat. Game board is rotated; child swats the card and carries out the learning cue.

Fortune teller. Child moves the card (in Fig. 4) back and forth to reveal the cue.

Fig. 4 Fortune teller template

Declarations

Conflict of interest No conflicts of interest to declare.

References

Ainsworth, S., Welbourne, S., & Hesketh, A. (2015). Lexical restructuring in preliterate children: Evidence from novel measures of phonological representation. *Applied Psycholinguistics, 37*(04), 997–1023. https://doi.org/10.1017/s0142716415000338

Bahrick, H. P., & Hall, L. K. (2005). The importance of retrieval failures to long-term retention: A meta-cognitive explanation of the spacing effect. *Journal of Memory and Language, 52*(4), 566–577. https://doi.org/10.1016/j.jml.2005.01.012

Beck, I. L., McKeown, M., & Kucan, L. (2013). *Bringing words to life: Robust vocabulary instruction*. Newyork: Guilford Press.

Biemiller, A., & Boote, C. (2006). An effective method for building meaning vocabulary in primary grades. *Journal of Educational Psychology, 98*(1), 44–62. https://doi.org/10.1037/0022-0663.98.1.44

Blachowicz, C. L., Fisher, P. J., Ogle, D., & Watts-Taffe, S. (2006). Vocabulary: Questions from the classroom. *Reading Research Quarterly, 41*(4), 524–539. https://doi.org/10.2307/1510288

Breadmore, H. L., Levesque, K., & Deacon, S. H. (2021). Special issue editorial: Advances in understanding the role of morphemes in literacy development. *Journal of Research in Reading, 44*(1), 1–9. https://doi.org/10.1111/1467-9817.12346

Carpenter, S. K., & DeLosh, E. L. (2005). Application of the testing and spacing effects to name learning. *Applied Cognitive Psychology, 19*(5), 619–636. https://doi.org/10.1002/acp.1101

CLPE (Centre for Literacy in Primary Education). (2018) Book lists. https://www.clpe.org.uk.

Coloma, C. J., De Barbieri, Z., Silva, M., & Rojas, D. (2022). Programa integrado de vocabulario y conciencia fonológica para preescolares con trastorno de desarrollo del lenguaje. *Onomázein, 57*, 241–256.

Damhuis, C., Segers, E., Scheltinga, F., & Verhoeven, L. (2016). Effects of individualized word retrieval in kindergarten vocabulary intervention. *School Effectiveness and School Improvement, 27*(3), 441–454. https://doi.org/10.1080/09243453.2015.1114500

Dickinson, D. K., McCabe, A., Anastasopoulos, L., Peisner-Feinberg, E. S., & Poe, M. D. (2003). The comprehensive language approach to early literacy: The inter-relationships among vocabulary, phonological sensitivity and print knowledge among preschool-aged children. *Journal of Educational Psychology, 95*(3), 465–481. https://doi.org/10.1037/0022-0663.95.3.465

Droop, W., Peters, S., Aarnoutse, C., & Verhoeven, L. (2005). Effecten van stimulering van beginnende geletterdheid in groep 2. *Pedagogische Studien, 82*, 160–180.

Duff, F. J., Reen, G., Plunkett, K., & Nation, K. (2015). Do infant vocabulary skills predict school-age language and literacy outcomes? *Journal of Child Psychology and Psychiatry, 56*(8), 848–856. https://doi.org/10.1111/jcpp.12378

Duke, N. K., & Cartwright, K. B. (2021). The science of reading progresses: Communicating advances beyond the Simple View of Reading. *Reading Research Quarterly, 56*(1), S25–S44. https://doi.org/10.1002/rrq.411

Dunn, L., Dunn, D., Styles, B., & Swells, J. (2009). *British Picture Vocabulary Scales* (3rd ed.). London: GL Assessment.

Ebbels, S. H., Bannister, L., Holland, B., & Campbell, L. (2022). Effectiveness of intervention focused on vocational course vocabulary in post-16 students with (developmental) language disorder. *International Journal of Language & Communication Disorders, 57*(6), 1334–1353. https://doi.org/10.1111/1460-6984.12758

Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly, 36*(3), 250–287. https://doi.org/10.1598/RRQ.36.3.2

Elleman, A. M., Lindo, E. J., Morphy, P., & Compton, D. L. (2009). The impact of vocabulary instruction on passage-level comprehension of school-age children: A meta-analysis. *Journal of Research on Educational Effectiveness*, *2*(1), 1–44. https://doi.org/10.1080/19345740802539200

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Sage.

Garlock, V. M., Walley, A. C., & Metsala, J. L. (2001). Age-of-acquisition, word frequency and neighborhood density effects on spoken word recognition by children and adults. *Journal of Memory and Language, 45*(3), 468–492. https://doi.org/10.1006/jmla.2000.2784

German, D. J., Schwanke, J. H., & Ravid, R. (2012). Word finding difficulties: Differentiated vocabulary instruction in the speech and language room. *Communication Disorders Quarterly, 33*(3), 146–156. https://doi.org/10.1177/1525740111405840

Gibbs, S., & Bodman, S. (2014). *Phonological Assessment Battery Primary* (2nd ed.). London: GL Assessment.

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

GOV.UK. (2015). *English indices of deprivation*. https://www.gov.uk/government/collections/english-indices-of-deprivation.

GOV.UK. (2018). OFSTED. https://www.gov.uk/government/organisations/ofsted.

Harm, M. W., & 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.

Jaccard, J. (1998). *Interaction effects in factorial analysis of variance*. Sage.

Janssen, C., Segers, E., McQueen, J. M., & Verhoeven, L. (2018). Comparing effects of instruction on word meaning and word form on early literacy abilities in kindergarten. *Early Education and Development, 30*(3), 375–399. https://doi.org/10.1080/10409289.2018.1547563

Kucera, H., & Frances, W. (1967). *Computational analysis of present-day American English*. Brown University Press.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods, 44*(4), 978–990. https://doi.org/10.3758/s13428-012-0210-4

Lawson-Adams, J., & Dickinson, D. K. (2020). Building lexical representations with nonverbal supports. *Reading Research Quarterly, 56*(3), 603–622. https://doi.org/10.1002/rrq.326

Lee, J. (2011). Size matters: Early vocabulary as a predictor of language and literacy competence. *Applied Psycholinguistics, 32*(1), 69–92. https://doi.org/10.1017/s0142716410000299

Lowe, H., Henry, L., & Joffe, V. (2019). The effectiveness of classroom vocabulary intervention for adolescents with language disorder. *Journal of Speech Language and Hearing Research, 62*(8), 2829–2846. https://doi.org/10.1044/2019_JSLHR-L-18-0337

Marulis, L. M., & Neuman, S. B. (2010). The effects of vocabulary intervention on young children's word learning: A meta-analysis. *Review of Educational Research, 80*(3), 300–335. https://doi.org/10.3102/0034654310377087

McGregor, K. K., Van Horne, A. O., Curran, M., Cook, S. W., & Cole, R. (2021). The challenge of rich vocabulary instruction for children with developmental language disorder. *Language, Speech & Hearing Services in Schools, 52*(2), 467–484. https://doi.org/10.1044/2020_LSHSS-20-00110

Metsala, J. L., & Walley, A. C. (1998). Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. In J. Metsala & L. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 89–120). Lawrence Erlbaum Associates.

Moran, E., & Moir, J. (2018). Closing the vocabulary gap in early years: Is' Word Aware' a possible approach? *Educational and Child Psychology, 35*(1), 51–64.

Munro, N., Lee, K., & Baker, E. (2008). Building vocabulary knowledge and phonological awareness skills in children with specific language impairment through hybrid language intervention: A feasibility study. *International Journal of Language and Communication Disorders, 43*(6), 662–682. https://doi.org/10.1080/13682820701806308

Nation, K. (2019). Children's reading difficulties, language, and reflections on the simple view of reading. *Australian Journal of Learning Difficulties, 24*(1), 47–73. https://doi.org/10.1080/19404158.2019.1609272

NFER (National Foundation for Educational Research). (2013). *Evaluation of the Phonics Screening Check: First interim report. May 2013*. Department for Education.

Noble, C., Cameron-Faulkner, T., Jessop, A., Coates, A., Sawyer, H., Taylor-Ims, R., & Rowland, C. F. (2020). The impact of interactive shared book reading on children's language skills: A randomized controlled trial. *Journal of Speech, Language, and Hearing Research, 63*(6), 1878–1897. https://doi.org/10.1044/2020_JSLHR-19-00288

Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology, 98*(3), 554–566. https://doi.org/10.1037/0022-0663.98.3.554

Ouellette, G., & Fraser, J. R. (2009). What exactly is a yait anyway: The role of semantics in orthographic learning. *Journal of Experimental Child Psychology, 104*(2), 239–251. https://doi.org/10.1016/j.jecp.2009.05.001

Parsons, S., & Branagan, A. (2014). *WordAware: Teaching vocabulary across the day, across the curriculum*. Milton Keynes: Speechmark.

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In Ludo Verhoeven, Carsten Elbro, & Pieter Reitsma (Eds.), *Precursors of functional literacy* (pp. 189–213). Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/swll.11.14per

Ricketts, J., Nation, K., & Bishop, D. V. M. (2007). Vocabulary is important for some, but not all reading skills. *Scientific Studies of Reading, 11*(3), 235–257. https://doi.org/10.1080/10888430701344306

Roodenrys, S., & Hinton, M. (2002). Sublexical or lexical effects on serial recall of nonwords? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(1), 29–33. https://doi.org/10.1037/0278-7393.28.1.29

Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical Evaluation of Language Fundamentals* (4th ed.). Agra: The Psychological Corporation.

Sénéchal, M., Ouellette, G., & Rodney, D. (2006). The misunderstood giant: On the predictive role of early vocabulary to future reading. In D. Dickinson & S. Neuman (Eds.), *Handbook of early literacy research* (pp. 173–182). London: Guilford.

Silverman, R. (2007). A comparison of three methods of vocabulary instruction during readalouds in kindergarten. *The Elementary School Journal, 108*(2), 97–113. https://doi.org/10.1086/525549

Snowling, M. J., & Hulme, C. (2020). Annual research review: Reading disorders revisited – the critical importance of oral language. *Journal of Child Psychology and Psychiatry, 62*(5), 635–653. https://doi.org/10.1111/jcpp.13324

Speech and Language UK (2023) *Listening to unheard children*. Speech and Language UK

Spencer, S., Clegg, J., Lowe, H., & Stackhouse, J. (2017). Increasing adolescents' depth of understanding of cross-curriculum words: An intervention study: Evaluating adolescent cross-curriculum vocabulary intervention. *International Journal of Language and Communication Disorders, 52*(5), 652–668. https://doi.org/10.1111/1460-6984.12309

St John, P., & Vance, M. (2014). Evaluation of a principled approach to vocabulary learning in mainstream classes. *Child Language Teaching and Therapy, 30*(3), 255–271. https://doi.org/10.1177/0265659013516474

Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology, 38*(6), 934. https://doi.org/10.1037/0012-1649.38.6.934

Storkel, H., & Morrisette, M. (2002). The lexicon and phonology: Interactions in language acquisition. *Language, Speech and Hearing Services in Schools, 33*(1), 24–36. https://doi.org/10.1044/0161-1461(2002/003

Vitevitch, M. S., Luce, P. A., Pisoni, D. B., & Auer, E. T. (1999). Phonotactics, neighborhood activation, and lexical access for spoken words. *Brain and Language, 68*(1–2), 306–311. https://doi.org/10.1006/brln.1999.2116

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., Hecht, S. A., Barker, T. A., Burgess, S. R., Donahue, J., & Garon, T. (1997). Changing relations between phonological processing abilities and word-level reading as children develop from beginning to skilled readers: A 5-year longitudinal study. *Developmental Psychology, 33*(3), 468–479. https://doi.org/10.1037/0012-1649.33.3.468

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (2013). *Comprehensive test of phonological processing* (2nd ed.). Austin: Pro-Ed.

Wegener, S., Beyersmann, E., Wang, H. C., & Castles, A. (2022). Oral vocabulary knowledge and learning to read new words: A theoretical review. *Australian Journal of Learning Difficulties, 27*(2), 253–278. https://doi.org/10.1080/19404158.2022.2097717

Widgit. (2007). *Widgit multimedia.* www.widgit.com.

Wordsmyth. (2014). *Children's dictionary*. http://www.wordsmyth.net.

Yopp, H. (1988). The validity and reliability of phonemic awareness tests. *Reading Research Quarterly, 23*(2), 159–177. https://doi.org/10.2307/747800