The effectiveness of classroom vocabulary intervention for adolescents with language disorder.

Corresponding author:

Hilary Lowe

Speech and Language Therapist and Honorary Research Fellow

Division of Language and Communication Science

School of Health Sciences

City, University of London

Northampton Square

London, EC1V 0HB, UK

Hilary.lowe@city.ac.uk

Tel: 07908 200778

Co-authors

Lucy Henry

Division of Language and Communication Science

School of Health Sciences

City, University of London

Northampton Square

London, EC1V 0HB, UK

Lucy.henry.1@city.ac.uk

Victoria L. Joffe

Professor, Enhancement of Child and Adolescent Language and Learning

Division of Language and Communication Science

School of Health Sciences

City, University of London

Northampton Square

London, EC1V 0HB, UK

v.joffe@city.ac.uk

Keywords:

vocabulary, adolescence, intervention, language disorder

Funding

This study was funded by a PhD studentship from the School of Health Sciences, City, University of London.

Declaration of interest

The authors report no conflicts of interest.

Abstract

Purpose

Phonological-semantic intervention has been shown to be effective in enhancing the vocabulary skills of children with language disorder in small-group or individual settings. Less is known about vocabulary interventions for adolescents with language disorder in whole-class models of delivery. The current study investigated the effectiveness of phonological-semantic vocabulary intervention for adolescents with language disorder, delivered by secondary school teachers within science lessons.

Methods

Seventy-eight adolescents with language disorder, aged 11 – 13 years, were taught science curriculum words by teachers in class, under two conditions: 1) 10 words taught through usual teaching practice; and 2) 10 matched words taught using an experimental intervention known as Word Discovery, which embedded phonological-semantic activities into the teaching of the syllabus. Ten similar control words received no intervention. Word knowledge was assessed pre-intervention, post-intervention, and follow-up.

Results

At pre-intervention, measures of depth of word knowledge and expressive word use did not differ between usual teaching practice and experimental words. At post-intervention, depth of knowledge of experimental words was significantly greater than that of usual teaching practice words. This significant advantage was not maintained at follow-up, although depth of knowledge for experimental words remained significantly higher at follow-up than at pre-intervention. At post-intervention, expressive use of experimental words was significantly greater than that of usual teaching practice words, and this significant difference was maintained at follow-up. There was no change in students' depth of knowledge or expressive use of no-intervention words over time, confirming that the findings were not due to maturity or practice effects.

Conclusion

The experimental intervention was more effective than usual teaching practice in increasing the word knowledge of participants. Clinical and teaching implications include the importance of intervening during the adolescent years, with classroom vocabulary intervention being a viable option for collaborative teacher and speech and language therapy/pathology practice.

Introduction

Vocabulary skills are often at risk for children and adolescents with language disorder (e.g. McGregor, Oleson, Bahnsen, & Duff, 2013). Vocabulary knowledge is a key predictor of reading comprehension, essential for academic progress (e.g. Nation & Snowling, 2004); and in the longer term, language disorder is associated with poorer outcomes in educational attainment, cognition, behaviour, social and emotional functioning and employment, well into adulthood (e.g. Johnson, Beitchman, & Brownlie, 2010). The term language disorder is used here to refer to difficulties with first language acquisition, which are likely to cause “a significant impact on social interactions or education progress” (Bishop, Snowling,Thompson, Greenhalgh, & CATALISE-2 consortium, 2017, p. 5). When citing previous research, the terminology of the authors is used. The current study investigates a new vocabulary intervention approach that incorporates a combination of phonological and semantic approaches, aimed at improving school-related vocabulary in adolescents with language disorder. The name given to the intervention is “Word Discovery.”

Phonological and semantic approaches to vocabulary intervention

The experimental Word Discovery intervention uses a combined phonological-semantic approach to enhance the vocabulary skills of children with language disorder, and is thus underpinned by theories of word learning (Bishop, 2014; Leonard, 1998; Stackhouse & Wells, 1997). A phonological-semantic approach links the phonological form (sound structure) with the semantic representation (meaning) of words; for example, the word *migration* begins with [m], has three syllables, and rhymes with *station* (phonological information); and it means when animals move to different areas of the world (semantic information). McGregor, Newman, Reilly, and Capone (2002) compared the object naming and semantic representations of 16 children with specific language impairment aged 5:0 –

7:11 with that of 16 age-matched typically developing (TD) peers, and found that the children with specific language impairment named fewer items and had sparser semantic representations, as assessed through a drawing task, than their age-matched peers. The value of adding phonological instruction to semantic instruction is particularly pertinent for children and adolescents who have language disorder, as they frequently have phonological as well as semantic weaknesses. Stackhouse and Wells (1997) and Lahey and Edwards (1999) have posited that weak phonological processing ability particularly affects naming, a task which requires production of the word. Kail and Leonard (1986) have, for example, acknowledged the role of phonological skills, and further proposed that naming is particularly dependent upon efficient semantic storage of words when they are being learned. Kail and Leonard (1986) argue that if words are inefficiently stored within the semantic system, this limits depth of word knowledge, and affects receptive vocabulary ability as well as a child's ability to retrieve a word in order to use it expressively. Hence, combining phonological and semantic strategies was the approach adopted in the current intervention.

There is some emergent evidence to support the effectiveness of phonological-semantic intervention in the secondary school age group (11 – 16 years of age). A systematic review of the vocabulary intervention literature with this age group (Lowe, Henry, Müller & Joffe, 2017) revealed tentative evidence for the impact of combined phonological-semantic intervention on receptive vocabulary (Murphy et al., 2017), depth of word knowledge (Spencer, Clegg, Lowe & Stackhouse, 2017), and expressive vocabulary (Ebbels et al., 2012).

For example, Murphy et al. (2017) explored the delivery of an adapted Vocabulary Enrichment Intervention Programme (VEIP: Joffe, 2011) in a randomised delayed intervention study, enhancing vocabulary skills through phonological and semantic intervention for curriculum words and developing independent word-learning skills. The adapted VEIP was delivered during 12 – 16 English lessons to 203 students aged 11:11 –

13:11 attending schools in Ireland in areas of social disadvantage. Significant improvement on standardised scores was reported for both the experimental group and the waiting control group on the Word Classes (Receptive), and Word Definitions subtests of the Clinical Evaluation of Language Fundamentals, Fourth Edition, UK (CELF-4 UK: Semel, Wiig, & Secord, 2006), so improvements on these two measures could not be accounted for by the intervention. However, the experimental group also made significant progress on the CELF-4 UK Word Classes (Expressive) subtest and on the British Picture Vocabulary Scale third edition (BPVS-3: Dunn, Dunn, Sewell, & Styles, 2011), a finding which was repeated in the waiting control group following their delayed intervention.   Given that progress on these two assessments during the control group's baseline period had been non-significant, the findings provided some evidence that the intervention had an impact on semantic relations and receptive vocabulary knowledge, although overall the results were mixed.

The study by Spencer et al. (2017) provides some further support for the effectiveness of vocabulary intervention on depth of word knowledge within mainstream secondary schools. These researchers worked with 35 12 – 13-year-olds who had low receptive vocabulary levels, in a mainstream secondary school in an area of social disadvantage in the UK, using a matched-groups delayed intervention design. Phonological-semantic intervention was carried out by speech and language therapists (SLTs) in small groups, one hour a week for an average of seven sessions, for ten cross-curricular verbs, e.g. *evaluate*. On a bespoke depth of word knowledge assessment, differences in progress on experimental word knowledge compared with control word knowledge for the intervention group were not significant. However, the waiting control group, following their delayed intervention, made significantly better progress with experimental word knowledge than with control word knowledge. Further, when the results of the two groups were combined, the increase in depth of word knowledge was significantly greater than zero for the experimental words, but not for the control words. Although these results were somewhat equivocal, they did provide some support for the effectiveness of phonological-semantic interventions.

Evidence for the effectiveness of a phonological-semantic approach in enhancing naming ability has also been provided in an individual intervention context (a specialist language setting) by Ebbels et al. (2012), who worked with 15 students aged 9:11 – 15:11, randomised to an intervention or waiting control group. The intervention was predominantly semantic, with a phonological element. Individual intervention was delivered in 15-minute sessions twice a week over a period of eight weeks. The authors reported significant progress for the experimental group, but not for the waiting control group, on the Test of Adolescent/Adult Word Finding (TAWF: German 1990), although not on the Test of Word Finding in Discourse (TWFD: German, 1991). Following their intervention, the waiting control group, assessed only on the TAWF, also made significant progress. Despite these promising results, the small sample size means that replication in larger-scale studies is needed.

Synthesising these sources of evidence suggests that although research is limited and findings are mixed, a phonological-semantic approach does have some potential to enhance the vocabulary knowledge of adolescents with language disorder.

Model of intervention delivery

Small group or individual models of intervention, necessitating withdrawal from the classroom, can have clinical, pedagogical, and practical disadvantages, particularly as students enter adolescence (Ehren, 2002). Furthermore, many children with language disorder are educated in mainstream schools but specialist speech and language support typically decreases as children move from primary to secondary education (Bercow, 2008; Ehren, 2002; Hollands, van Kraayenoord & McMahon, 2005; Lindsay, Dockrell, Mackie, & Letchford, 2005; Royal College of Speech and Language Therapists/ICAN, 2018). A universal model, whereby intervention takes place in a whole-class context, may therefore be particularly relevant for the secondary school setting.

Snow, Lawrence, and White (2009), implemented a class-based vocabulary intervention (Word Generation) for the duration of one academic year in the USA, for cross-curricular words with 11 – 14 year-olds. These students had low language levels in association with second language learning and social disadvantage. Word Generation included: encountering target words in semantically rich contexts within motivating texts; recurrent exposure to the words in varied contexts; using the word orally and in writing; explicit instruction in word meaning; and explicit instruction in independent word learning strategies (Snow et al., p.327). The participating 697 students made progress, relative to 319 controls, on a multiple-choice reading task involving 40 of the 120 intervention words. The experimental group learnt a mean of 4.43 words, while the control group learnt a mean of 1.95 words (d = 0.21, small effect size). Furthermore, vocabulary improvement was found to significantly predict scores on the Massachusetts Comprehensive Assessment System, a curriculum assessment used in the USA.

However, in a review of vocabulary instruction for older children, Ford-Connors and Paratore (2015) found little evidence of in-depth vocabulary teaching in schools with this age-group. This finding is corroborated by a survey of speech and language therapists/pathologists (SLT/Ps) and teachers working in mainstream secondary schools (Lowe, Henry, Wallinger, & Joffe, in preparation), which found that a phonological-semantic approach was widely used in speech and language therapy practice within the secondary school age group, but that a semantic and literacy-based approach was more likely to be used by mainstream secondary school teachers. Thus, adolescents with language disorder, for whom phonological-semantic input may be necessary, could be at a disadvantage in terms of vocabulary support in the classroom.

The only study we could identify investigating universal vocabulary intervention for adolescents with language disorder was that of Murphy et al. (2017). However, their study did not compare different models of delivery. Nevertheless, a study with younger children (Throneburg, Calvert, Sturm, Paramboukas, & Paul, 2000) did suggest that delivery of intervention in the classroom, facilitated by teacher / SLT/P collaboration, can be more effective for the vocabulary learning of young children with language disorder than a withdrawal model of intervention.

Therefore, the current study builds on a feasibility study by Lowe & Joffe (2017), which took phonological-semantic elements of intervention typically delivered in individual or small-group models, and applied them to a universal model, to be implemented by teachers, embedded within the science curriculum. The feasibility study employed a within-subjects design with a class of 15 students who had low vocabulary levels. The class teacher taught 10 science curriculum words using phonological-semantic activities, and 10 words using her usual teaching practice, which consisted of semantic and literacy-based activities such as word-picture matching. The outcome measure was a bespoke definition production task. The inclusion of some high-frequency words resulted in ceiling effects; however, once the highest frequency words were omitted from analysis, increase in knowledge for the five lowest frequency words was marginally significant in favour of the experimental condition. This word-learning approach and the service delivery model of intervention were positively received by participating students and their teacher.

The current study builds on these preliminary findings by assessing a phonological-semantic classroom vocabulary intervention approach (Word Discovery), comprising a package of evidence-based intervention techniques. The phonological activities included awareness and practice of initial sound, syllable, and rhyme in relation to the targeted experimental words. The semantic activities included awareness and practice of the semantic features of the words in terms of function, location, attribute and group. In addition, the intervention took a holistic perspective encompassing other factors critical to word learning in adolescence. These included: linking new words to prior knowledge and adding them to an existing lexicon (Dockrell, Braisby, & Best, 2007); accompanying speech with the written word (Ricketts, Dockrell, Patel, Charman, & Lindsay, 2015); and direct instruction on how to derive meaning from context, to develop independent word learning skills (Nash & Snowling, 2006).

Aims of the current study

The aim of the study was to examine the effectiveness of Word Discovery intervention in increasing participants' knowledge of science curriculum words. Participants were adolescents with language disorder. It was hypothesised: (1) that the increase in *depth of knowledge* of experimental words (taught using Word Discovery) would be greater than that for control words (taught through usual teaching practice), from pre- to post-intervention and from post-intervention to follow-up. It was further hypothesised: (2) that the increase in *expressive use* of experimental words (taught using Word Discovery) would be greater than that for control words (taught through usual teaching practice), from pre- to post-intervention and from post-intervention to follow-up.

Methods

Study design

The study employed a within-subjects repeated measures design. Levels of depth of knowledge and expressive use of experimental, usual teaching practice, and no-intervention words were compared at pre-intervention, post-intervention, and follow-up time-points for each condition. The study phases are depicted in Figure 1.

Word knowledge assessments	**Intervention**	**Word knowledge assessments**	Teacher training	**Intervention**	**Word knowledge assessments**	**Word knowledge assessments**
Pre- usual teaching practice	Topic 1: usual teaching practice	Post- usual teaching practice *Pre-experimental intervention*		*Topic 2: experimental classroom vocabulary intervention (Word Discovery)*	Follow-up usual teaching practice *Post-experimental intervention*	*Follow-up experimental intervention*

9 weeks 9 weeks 5 weeks

Figure 1. Overview of study phases

School characteristics

Ethical approval for the study was received from the relevant University ethics committee. Signed informed consent was firstly obtained from the head teacher and then from science teachers, parents, and students. Eight non-selective mainstream secondary schools (i.e. schools for 11 – 16 year-olds, to which admittance is not determined by academic ability) from a wide geographical and socio-economic spread in England took part in the study. Using eligibility for free school meals as a proxy for socio-economic status, the average amongst participating schools was 13.9%, consistent with the national average of 13.6% (Department for Education [DfE], 2018). Demographic characteristics of the schools, as at the time of the study, are included in Table 1.

Table 1. School characteristics

School	Age range†	Number on roll†	PAN	Gender	Ofsted rating††	Geo-graphical region of England	% of pupils eligible for free school meals††	Decile of Deprivation according to Index of Multiple Deprivation†††
1	11-18 (Y7,8)	242	120	boys	Not available*	Greater London	20.7	2
2	11-18 (Y7,8,9)	271	90	mixed	Good	South East	5.9	9
3	11-16	1,181	230	mixed	Outstanding	South East	8.3	8
4	11-18	1,205	210	mixed	Good	Greater London	22.8	6
5	3-19	2,524	210	mixed	Requires Improvement	North	31.0	9
6	11-18	1,200	180	mixed	Outstanding	Greater London	12.9	4
7	11-19	1,476	250	mixed	Good	East	7.7	8
8	11-18	1,513	210	mixed	Outstanding	Midlands	1.9	8

Key: PAN = Published Admission Number: the number of students admitted each year in Year 7 (from schools' own websites).

* As a recently opened school, Ofsted rating was not available at the time of the study. In May 2017, Ofsted rating was Good. Ofsted is a body within the UK government which inspects and regulate services providing education. Schools are rated on a scale of outstanding – good – requires improvement – inadequate.

† Obtained from Ofsted Inspection Reports (Ofsted, 2017).

†† Obtained from Edubase2 (DfE, 2017a). Children are eligible for free school meals if their parents are in receipt of welfare benefits.

††† Neighbourhoods in the first decile are amongst the 10% most deprived neighbourhoods in England, and neighbourhoods in the tenth decile are amongst the 10% least deprived (Department for Communities and Local Government, 2015).

Teacher participants

The inclusion criterion for teacher participants was that they would be teaching science to student participants throughout the timescale of the study. (Student participant inclusion criteria are stated in the next section). Thirty-four teachers were invited, of whom 30 took part; nine male and 21 female. The mean number of years' overall teaching experience was 7.2 years (range <1 – 25 years), with mean secondary school teaching experience of 7.4 years (range <1 – 25 years). Teachers had previously received on average one day's training in speech, language, and communication needs (range 0 – 4 days).

Student participants

To be eligible for recruitment, students were required to have a verbal standard score (SS) on a test of general intellectual ability of below 85, and a nonverbal SS of equal to or higher than the individual's verbal score, but not below 70. In seven schools, this information was obtained from students' performance on the Cognitive Attainment Test (CAT: GL Assessment, 2015), an online assessment frequently taken by students on entry to secondary school in the UK. This measure was used for recruitment because, due to the high comorbidity of spoken and written language difficulties (e.g. McArthur, Hogben, Edwards, Heath, & Mengler, 2000) it was deemed an appropriate way of identifying students with potential language disorder. In the remaining school, which did not utilise the CAT, the verbal SS was obtained from students' performance on the Access Reading Test (Crumpler & McCarty, 2006), a paper-based reading assessment used routinely by the school. The nonverbal SS was obtained from the Matrix Reasoning subtest of the Wechsler Abbreviated Intelligence Scale, second edition (WASI-2: Wechsler, 2011), administered by the first author because no school-administered nonverbal measures were available. Participants were also required to have lived in the UK for at least two years, to allow for the acquisition of functional proficiency in English (MacSwan & Pray [2005] report between 1.5 and 5 years to achieve parity with native speakers, depending on age of arrival).

To be eligible for inclusion, students were required to score at least one standard deviation (*SD*) below the mean on at least one of five language assessments administered at baseline or on the CAT verbal score (CATV). These inclusion criteria were chosen to identify students who would demonstrate language difficulties potentially sufficient to impede access to the curriculum. One hundred and three students were recruited to the study, but for the following reasons, 25 did not take part. One student was found to have a CATV SS greater than 85 and age-appropriate scores on all language assessments. Five students left school or changed classes during the study; one opted out; two were absent at assessment points; and there were 16 students whose teachers opted out or delivered the intervention with some of their classes but not others. Thus, there were data at all time points for 78 students, aged 11 – 14 years; 52 male and 26 female. This imbalance in gender was partly due to the tendency for language disorder to be more prevalent in boys than girls (Tomblin et al., 1997), and partly because school 1 was an all boys' school. Forty-one students were in Year 7, 29 were in Year 8, and eight were in Year 9 (equivalent to Grades 6, 7 and 8 in the USA). Mean chronological age was 12:3 (*SD* = 9 months: range 11:3 to 14:0). Twenty-eight (35.9%) of the student participants were eligible for Pupil Premium (DfE, 2016), an indicator of low socio-economic status. Ten participants (12.8%) had a medical condition not usually associated with language disorder (e.g. asthma, diabetes). Three students (3.9%) had conditions which may be associated with language disorder (Down Syndrome, foetal alcohol syndrome, and perforated eardrums). Twelve participants (15%) were in possession of a statement of educational need, or education, health and care plan (EHCP)[1], and thirty-seven participants (47%) were in receipt of school-based support without a statement or EHCP. Five participants (6.4%) were receiving speech and language therapy intervention. Of the whole cohort, 12 participants (15%) were in receipt of school-based and/or speech and language therapy vocabulary support in addition to the experimental intervention. A range of ethnic origins were represented in the cohort, but as ethnicity was reported in differing ways from school to school, it was not possible to amalgamate the data. Forty-nine participants (63%) were monolingual English speakers. The remaining 29 participants were bilingual or multilingual, with 14 of these reporting that English was the main language spoken at home. Seventy-seven students had been living in the UK for at least two years. The remaining student had lived in the UK for one year, and was retained in the study as she met all other criteria and was judged to have a functional proficiency in English. Although in some cases students were absent on the day that the words were introduced, they were present on other days when word-learning activities took place. Therefore, all students who had been present for any word-learning activity were counted as having taken part in the intervention.

Measures

Language and cognitive profiling

At baseline, students were assessed on the following assessments: the Vocabulary and Matrix Reasoning subtests of the WASI-2 (a definition production task and a nonverbal abilities

[1] Statements of educational need entitled schools in the UK to funding in order to meet individual needs. They were superseded by EHCPs in 2014.

task); the BPVS-3 (a multiple choice receptive oral vocabulary measure); the Recalling Sentences subtest of the CELF-4 UK (a sentence repetition task); the Listening Recall subtest of the Working Memory Test Battery for Children (WMTB-C: Gathercole & Pickering, 2001; a verbal working memory measure); and the Spoonerisms subtest[2] of the Phonological Awareness Battery (PhAB: Frederickson, Frith, & Reason, 1997; a measure of phonological awareness ability).

All assessments were administered individually by the first author in a quiet room in school, during school time, and scored according to the relevant examiner's manual. Standardised scores were derived for all standardised assessments, with a mean of 100 and a *SD* of 15. Table 2 shows the language and cognitive assessment scores of student participants. Ninety-one percent (71/78) scored -1 *SD* below the mean on at least two language measures, with 69% (54/78) scoring -1 *SD* below the mean on three or more language measures.

[2] In this task students are given two words and are asked to exchange the first two phonemes of each word to form a nonsense phrase e.g. *King John* becomes *Jing Kon.*

Table 2. Language and cognitive profiles of student participants

Assessment (N=78 except where stated)	**Mean SS (*SD*)**	**Minimum**	**Maximum**	**Number (%) with SS <85**
CATV *	77.96 (6.98)	59	104	76 (97.4%)†
CATNV (N=70) **	88.31 (8.53)	73	111	27 (34.6%)
WASI-2 Vocabulary	88.46 (8.87)	67	104	23 (29.5%)
WASI-2 Matrix Reasoning	92.05 (10.39)	64	121	18 (23.1%)
BPVS-3 Receptive Vocabulary	79.19 (9.20)	69	105	60 (76.9%)
CELF-4 UK Recalling Sentences	79.53 (14.44)	56	110	45 (57.7%)
WMTB-C Listening Recall	88.36 (17.51)	57	122	28 (35.9%)
PhAB Spoonerisms	89.03 (8.42)	69	117	20 (25.6%)

* Data represents Access Reading Test SS instead of CATV SS for participants from school 8.

**No school-administered nonverbal measure was available for participants from school 8

† After recruitment, two students were found to have a CATV SS of >85 (SS 93 and 104 respectively), but as they demonstrated difficulties on at least one of the language profiling assessments, they were retained in the study.

Key: CATV = Cognitive Ability Test Verbal subtest (GL Assessment, 2015)

CATNV = Cognitive Ability Test Nonverbal subtest (GL Assessment, 2015)

WASI-2 = Wechsler Abbreviated Scale of Intelligence, second edition (Wechsler, 2011)

BPVS-3 = British Picture Vocabulary Scale (Dunn et al., 2011)

CELF-4 UK = Clinical Evaluation of Language Fundamentals, 4th edition, UK (Semel et al., 2006)

WMTB-C = Working Memory Test Battery for Children (Gathercole & Pickering, 2001)

PhAB = Phonological Assessment Battery (Frederickson et al., 1997)

Word knowledge assessments

The primary outcome measures in the current study pertained to depth of word knowledge and expressive word use of subject-specific words from the curriculum syllabus. The words were chosen from science, because science is a core (compulsory) subject in the UK secondary school curriculum (DfE, 2014), and is noted for its high content of subject-specific vocabulary, much of which is abstract or technical (Woodward & Noell, 1991). Science vocabulary has been found to be challenging for adolescents with language disorder (Forwood, 2014).

The head of science in each school supplied a list of key subject-specific words from two topics that would be taught sequentially during the timeframe of the study, and the first author sourced no-intervention control words from future science syllabi. From these, three lists of words were created (active control, experimental, passive control) which were matched as closely as possible for (1) phonological complexity, (2) concreteness and (3) frequency according to the Zipf scale (van Heuven, Mandera, Keuleers, & Brysbaert, 2014).

Each student was assessed on one set of 30 words as follows: 10 active control words from topic 1: to be taught through usual teaching practice; 10 experimental words from topic 2: to be taught using the experimental intervention, Word Discovery; and 10 passive control words: words from future science topics which were not taught during the timescale of the study. Because student participants were taught in 46 separate classes, there were 22 different sets of words in total. Appendix A contains information on phonological complexity, concreteness, and frequency for one set of words as an example.

As the intervention targeted sets of curriculum words, a bespoke non-standardised tool to measure increases in word knowledge was devised consisting of a definition production task, in which the participants were required to describe the meaning of each word. The first author administered all assessments according to a flow chart protocol, and all assessments were audio-recorded for later transcription. Students were given a visual prompt card containing squares coloured red, amber, green, and a green star, with the purpose of engaging students in the assessment, and to draw out their maximum knowledge about each word. An explanation of the task and an example were given. Each written word was then read out by the assessor and shown to the student one by one, and the assessor asked the student "What does ….. mean?" Dependent on the responses made by the participants, staged prompts were given by the assessor. These prompts included:

Can you tell me anything about what it means?

Can you tell me more exactly what it means in science?

Can you think of anything else it means in science?

Can you use the word in a sentence?

At baseline, responses were scored according to definitions provided by the science teacher. Responses were collated along with the rating they had been awarded, generating a scoring guideline sheet for each set of words, so that marking was consistent across participants and across time-points.

A scale to describe levels of word knowledge, first proposed by Dale (1965) and adapted in several previous studies (e.g. McGregor et al., 2013), was used. In the current study, the word knowledge measure consisted of two scales, in order to provide distinct information regarding, firstly, how well the student knew the meaning of the word and, secondly, how well they could use the word: both important skills for access to the curriculum and examination success.

1) *Depth of word knowledge*, measured using a definition production task, primarily assessing semantic representation.

2) *Expressive word use*, assessed by asking the participant to use the word in a meaningful sentence. This gave additional insight into semantic representation, as well as phonological representation.

The scales and scoring system are detailed in Table 3. An expressive word use score could only be given if the student scored the maximum (score = 2) on the depth of word knowledge scale. If that criterion was met, a score of 1 on the expressive word use scale was awarded if the student produced the word with phonological accuracy in a meaningful sentence. Responses had to meet criteria for sentence structure and content (taken from the CELF- 4 UK Formulated Sentences subtest)[3] as well as speech production (taken from the Expressive One Word Picture Vocabulary Test [Brownell, 2010])[4].

The validity of the depth of word knowledge assessment was measured by correlating the baseline depth of word knowledge assessment scores with the students' scores on the WASI-2 Vocabulary subtest, which is also a definition production task. There was a significant positive correlation between the depth of word knowledge scores and WASI-2 Vocabulary raw scores (Pearson's $r = .492$, $p < .01$).

[3] i.e. a meaningful sentence with no more than two deviations in syntax or semantics' (Semel et al., 2006, p.33).
[4] i.e.. dropping, substituting, adding, or transposing a sound or syllable was counted as an error. If a sound was not within a student's phonetic inventory, habitual pronunciations were counted as correct e.g. if unable to produce [ʃ] (sh), [inhəleɪsən] ("inhalasun") would be counted as correct for *inhalation*.

Table 3. Word knowledge assessment scoring system

Rating		**Score on depth of word knowledge scale**	**Score on expressive word use scale**
Red	Student does not demonstrate any knowledge of word meaning	0	
Amber	Student indicates some, but imprecise, understanding of word meaning	1	
Green	Student demonstrates clear understanding of meaning in the science context	2	
Green star	Student can use the word in a spoken sentence		1
	Maximum achievable for 10 words	20	10

In order to establish reliability of the assessment scoring, an SLT not otherwise connected with the study was trained by the first author in the use of the word knowledge assessment and its scoring. Sample audio-recordings of the study cohort data were scored together to train the SLT in the application of the scoring guidelines. This SLT then second-marked 25% of the depth of word knowledge and expressive word use assessments at all time-points directly from the audio-recordings. The SLT was blind to the status of the words, and blind to the marking of the first author. Unweighted Cohen's kappa, computed online (Lowry, 2001-2017), indicated strong agreement between the two raters, $\kappa = .841$ (95% CI, .820 to .861), suggesting that the scoring was reliable.

Procedure

Usual teaching practice strategies

Usual teaching practice data were gathered through topic 1 strategy records completed by the teachers, and through lesson observations by the first author. (See Tables 4 and 5).

Table 4. The most frequently used strategies reported by teachers

Number of teachers (out of 27)	**Number of instances**	**Strategy**
12	22	Definition games
10	11	Spelling
9	44	Practical demonstration / experiment
8	20	Give definitions
8	11	Display key words with a visual image
7	19	Give examples
6	10	Discussion
6	8	List key words on the board
6	8	Teach how to derive meaning from morphology
6	7	Encourage students to draw on personal experience related to the word through scaffolded questioning
6	7	Students write the word
6	6	Use of a visual image - video
5	7	Students say the word aloud
5	6	Students to generate their own definition
5	5	Reading
5	5	Students say the word in a sentence
< 5	1 - 6	Use of a visual image - diagram Repetition Teach phonological awareness of the word Students write the word in a sentence Semantic feature analysis Students write the word - cloze Students write the definition Encourage students to think of a personal experience related to the word Praise Give synonyms Word generation Find associated words Develop student awareness by encouraging students to identify unknown words Make word vocally salient Students self-rate their own knowledge

Table 5. The most frequently used strategies noted in lesson observations

Number of teachers (out of 14)	Number of instances	Strategy
8	10	List key words on the board
7	8	Give definitions
6	6	Give definition - paraphrase
< 5	1 - 5	Repetition Reading Students write the word in a cloze activity Encourage students to draw on personal experience related to the word through scaffolded questioning Students say the word aloud Give synonyms Signpost Students write the definition Definition games Teach how to derive meaning from morphology Teacher elicits specific word Students say the word in a sentence Students write the word Use of a visual image Word generation Map word to object Give examples of use in multiple contexts Give examples Make word vocally salient Teach phonological awareness of the word Students self-rate their own knowledge of the word Use of a visual image - diagram Song

Only two teachers (eight instances) reported the use of specific aspects of semantic feature analysis such as drawing attention to function, location, attribute or group, and only three teachers (three instances) reported using activities which involved phonological awareness. These data indicated that vocabulary teaching strategies used by teachers predominantly took a semantic (but not specifically semantic feature analysis) and literacy perspective.

Teacher training

The experimental word-learning intervention activities were taught to participating teachers in each school, in a one-hour interactive training session, led by the first author. The training session took place between topic 1 (usual teaching practice) and topic 2 (Word Discovery). All resources necessary for the intervention activities and record-keeping were supplied to teachers both in hard copy and electronically. Teachers were asked to deliver the Word Discovery activities within each science lesson for the duration of topic 2. A suggested schedule for implementing the word-learning activities over 10 lessons was provided for teachers.

The experimental intervention

The intervention consisted of seven components, as follows.

Self-rating checklist. At the beginning of topic 2, students were given a self-rating checklist. The 10 experimental words were listed on one sheet, against three columns headed with a sad face (representing no knowledge of word meaning), a non-committal face (representing some knowledge), and a smiley face (representing secure knowledge). Teachers were asked to read the words aloud to the students, who then rated their own knowledge of the words individually by ticking the appropriate column. The self-rating checklist was done once at the beginning of the topic, and once at the end of the topic so that students could review their own learning.

Visual image displayed with written word. An image representing each experimental word along with the written word, each on an A4 (210 × 297 mm) laminated sheet, was supplied to the teachers for display in class throughout the topic.

Word detective. Words were introduced in context by reading aloud a piece of text from a lesson presentation on PowerPoint (Microsoft, 2016), and the teacher modelled what to do when encountering a new word. The concept of being a word detective was taken from Joffe (2011), and the word detective prompt card was devised as a mnemonic to remind the students of four key strategies for finding out the meaning of a new word (to look for morphological clues in the structure of the word; to look for contextual clues in the sentence or paragraph containing the word; asking another person; and using a dictionary). Teachers were asked to model being a word detective for at least three of the experimental words.

Word map. A word map (based on Elks & McLachlan, 2008) was used to introduce new concepts, forming a framework for exploring the meaning of the words. Out of all the activities, the word map was intended to be the one where the majority of the teaching of new curriculum content would occur. A word was written in the centre of the word map. On one side of the word, lines led to spaces in which to write: the number of syllables; the initial phoneme; and words which rhyme with or sound like the word. This latter space also allowed for discussion about morphology and linking with other similar words through examining the root, prefix, and suffix. On the other side, lines led to spaces in which to write or draw: the function of the object; its location; its constituent parts; what category it belonged to; and something that personalised the word to the student's own experience. Drawing was used as much as possible to provide visual support, and to allow those with literacy difficulties to demonstrate their knowledge. The teachers initially did the word maps on the board as a whole-class activity, but once the students were familiar with the word map framework, it was used flexibly, for example on a printed sheet in pairs, individually, or as homework. Teachers were asked to do a word map for at least five of the experimental words.

Word wise quickie (Elks & McLachlan 2008). This is a short verbal activity in which students are given a word: they think of a meaning, think of a sound (i.e. the number of syllables, initial phoneme, or a rhyme), and use the word in a spoken sentence. Teachers were asked to do a word wise quickie for at least the five words which had not been explained with a word map. It could be done as a whole class or in pairs. A prompt card was provided for teachers to use as a mnemonic for themselves or to display.

Sound and meaning bingo. The experimental words were written on the board and students each chose a given number of them to write in a grid. The teacher gave a sound and meaning clue for a word, and students put their hand up if they had this word in their grid. Examples of clues for *kinetic* might be: “It begins with k and means movement energy” or “It rhymes with frenetic and is the type of energy created by a rolling ball”. One student with their hand up was asked to say the word aloud, and students who had it in their grid crossed it off. Play continued until one of the students had crossed off all their words and called bingo. Teachers were asked to play sound and meaning bingo three times in all, towards the end of the topic.

Key word sheet. This contained 10 boxes and the alphabet down the centre. To complete an entry in the key word sheet, the student carried out the following tasks: writing the word in a box; placing a dot under each syllable; drawing or writing their own understanding of what the word meant; and drawing a line to link it with its initial letter. The key word sheet was placed in the student’s book or folder at the beginning of the topic for easy access. The teachers were asked to provide opportunity for the students to do a key word sheet entry for all 10 experimental words.

Fidelity measures

Fidelity to the intervention protocol by participating teachers was measured in three ways.

Teachers’ records. Topic 2 strategy records gave the researcher information on: how many of the word-learning activities had been done; on what date; and with which words. Twenty-eight topic 2 strategy records were received, from 20/30 teachers. For four teachers who did not return their strategy records, information was gained verbally or via email.

Students' work. At the end of topic 2, the researcher collected photocopies of relevant work produced by participating students. In some cases, students' work was not available to the researcher; for instance, if the students' books were at home for revision. Work was obtained from 63 students: word maps for 53 students; self-rating checklists for 46 students; key word sheets for 36 students; bingo sheets for 11 students; and word wise quickies for 3 students. These were anonymised upon receipt.

Lesson observations. Twenty lessons (17 teachers) were observed during topic 2. The researcher collected data on: how new words were taught; frequency of exposure of experimental and control words; and duration of the word-learning activities. The length of time each word-learning activity took was recorded in order to calculate an average length of time for each word-learning activity.

Following topic 2, data from these three sources of information were cross-referenced and collated to gain an overall picture of the intervention which each participant received. Out of the 46 classes, there was evidence that 23 utilised the self-rating checklist at least once, at least 18 displayed visual images,19 modelled being a word detective, 32 used word maps, 20 used word wise quickies, 22 played sound and meaning bingo, and 25 completed key word sheets.

Dosage

From the fidelity data (teachers' records, students' work, and lesson observations), it was calculated that the total amount of time each class spent on Word Discovery activities in topic 2 ranged between 6.5 and 135.5 minutes (average 62.5 minutes).

To avoid contamination of the data, teachers were not asked during topic 1 to record the amount of time spent on teaching words. The relative time spent in each condition was therefore made by comparing the total amount of time spent in lessons for each topic, and by examining word exposure. Topic 1 was delivered in an average of 11.6 lessons (range 5 – 20) over an average period of 4.25 weeks (range 2 - 9). Topic 2 was delivered in an average of 13.1 lessons (range 6 – 27), over an average period of 4.33 weeks (range 2 - 9). A related samples *t*-test showed that the difference in the number of lessons between topic 1 and topic 2 was not significant (t (35) = -1.542, p = .132). All lessons were between 50 and 60 minutes long.

Word exposure

In their strategy records for both topic 1 and topic 2, teachers recorded which words were taught, and estimated how often they spoke each word to the class. The mean number of words taught in topic 1 was 8.5 out of 10, and in topic 2, it was 8.8 out of 10. Wilcoxon's signed ranks test, employed because the data were not normally distributed, indicated that this difference was not significant (Z = -1.593, p = .111). There was a marginally significant difference between the conditions in terms of the amount of exposure the words received (Z = -1.965, p = .049), with the experimental words receiving more exposure (M = 9.7; range per class 5.1 – 19.9) than the usual teaching practice words (M = 8.3; range per class 1.6 – 23.4). Teachers did not have access to the no-intervention words; exposure of these, therefore, was measured only by researcher lesson observations. None of the no-intervention words were observed to have been used, in either phase of the study.

Results

Depth of word knowledge

Depth of word knowledge data were analysed using SPSS 23 (IBM Corp, 2015). Means (*M*) and standard deviations (*SD*) for depth of word knowledge scores in each condition and at each time point are presented in Table 6.

Table 6. Mean scores, ranges, and confidence intervals for depth of word knowledge in each condition and at each time point

	Pre-intervention **M *(SD)*** ***(range)*** ***(95% Confidence Interval)***	**Post-intervention** **M *(SD)*** ***(range)*** ***(95% Confidence Interval)***	**Follow-up** **M *(SD)*** ***(range)*** ***(95% Confidence Interval)***
Usual teaching practice words out of 20	4.14 (2.75) (0 – 11) CI 3.53 to 4.75	5.72 (3.29) (0 – 15) CI 4.88 to 6.56	5.38 (3.36) (0 – 14) CI 4.63 to 6.13
Experimental words (Word Discovery) out of 20	3.50 (2.51) (0 – 10) CI 2.94 to 4.06	6.96 (3.87) (0 – 17) CI 6.1 to 7.82	6.17 (3.80) (0 – 16) CI 5.33 to 7.01
No-intervention words out of 20	.92 (1.27) (0 – 5) CI 0.68 to 1.2	.99 (1.47) (0 – 8) CI 0.66 to 1.3	.90 (1.37) (0 – 8) CI 0.6 to 1.2

To investigate depth of word knowledge performance in each experimental condition, a 3 (Condition: usual teaching practice, experimental (Word Discovery), no intervention) x 3 (Time: pre, post and follow up test) related (repeated measures) ANOVA was conducted, followed by planned pairwise comparisons with Bonferroni corrections. Where Mauchly's test indicated that the assumption of sphericity had been violated, a Greenhouse-Geisser correction was employed. There was a significant main effect of time, $F(2, 154) = 74.040$, p

$< .001$, $\eta_p^2 = .490$, large effect size, and a significant main effect of condition (sphericity not assumed), $F(1.968, 151.545)$, = 137.872, $p < .001$, $\eta_p^2 = .642$, large effect size. Importantly, and as predicted, there was a significant Time x Condition interaction effect (sphericity not assumed), $F(2.643, 203.516) = 26.080$, $p < .001$; $\eta_p^2 = .253$, large effect size, observed power 1.0.

To explore the interaction, planned comparisons with Bonferroni corrections were conducted to compare the effects of time-point for usual teaching practice, experimental (Word Discovery) and no intervention words separately. Depth of word knowledge of *usual teaching practice* words was significantly greater at the post-intervention point (*M* = 5.72, *SD* = 3.29) than at pre- intervention (*M* = 4.14, *SD* = 2.75), $p < .001$, $d = 0.52$, medium effect size; and there was no significant change between post-intervention and follow-up (*M* = 5.38, *SD* = 3.36), $p = .272$. This indicated that students' depth of word knowledge of usual teaching practice words increased significantly following usual teaching practice, and that this increase was maintained five weeks later. This was confirmed by a significant difference between pre-intervention and follow-up scores ($p < .001$). Depth of word knowledge of *experimental* words was significantly greater at the post-intervention point (*M* = 6.96, *SD* = 3.85) than at pre-intervention (*M* = 3.50, *SD* = 2.51), $p < .001$, $d = 1.09$, large effect size. Depth of word knowledge at follow-up (*M* = 6.17, *SD* = 3.80) was significantly lower than at post-intervention ($p = .002$), but still significantly greater than at pre-intervention ($p < .001$). This indicated that students' depth of word knowledge of experimental words increased significantly following the experimental intervention, and that this increase was partially maintained five weeks later. There was no significant change in depth of word knowledge of no-intervention words between pre-intervention (*M* = 0.92, *SD* = 1.27) and post-intervention (*M* = 0.99, *SD* =1.47), $p = 1.000$; or between post-intervention and follow-up (*M* = 0.90, *SD* =1.37), $p = 1.000$.

Further planned comparisons with Bonferroni corrections examined the effects of condition at each time-point. At the *pre-intervention point*, depth of word knowledge of usual teaching practice words (*M* = 4.14, *SD* = 2.75) was numerically greater than that of the experimental words (*M* = 3.50, *SD* = 2.51), but this difference was not significant ($p = .137$). Depth of word knowledge of no-intervention words (*M* = 0.92, *SD* = 1.27) was significantly lower than that of both usual teaching practice words ($p < .001$) and experimental words ($p < .001$). At the *post-intervention point*, depth of word knowledge of experimental words (*M* = 6.96, *SD* = 3.85) was significantly greater than that of usual teaching practice words (*M* = 5.72, *SD* = 3.29), $p = .015$. Depth of word knowledge of no-intervention words (*M* = 0.99, *SD* = 1.47) was significantly lower than that of both usual teaching practice words ($p < .001$) and experimental words ($p < .001$). At the *follow-up point*, depth of word knowledge of experimental words (*M* = 6.17, *SD* = 3.80) was still numerically greater than that of the usual teaching practice words (*M* = 5.38, *SD* = 3.36), but this difference was not significant ($p = .224$). Depth of word knowledge of no-intervention words (*M* = 0.90, *SD* = 1.37) was significantly lower than that of both usual teaching practice words ($p < .001$) and experimental words ($p < .001$).

Expressive word use

Means and standard deviations for expressive word use in each condition and at each time point are presented in Table 7.

Table 7. Mean scores, ranges, and confidence intervals for expressive word use in each condition and at each time point

	Pre-intervention M *(SD)* *(range)* (95% Confidence Interval)	**Post-intervention M *(SD)* *(range)* (95% Confidence Interval)**	**Follow-up M *(SD)* *(range)* (95% Confidence Interval)**
Usual teaching practice words out of 10	.58 (.91) (0 - 4) CI 0.39 to 0.78	.96 (1.39) (0 - 6) CI 0.65 to 1.27	.97 (1.37) (0 - 5) CI 0.66 to 1.27
Experimental words (Word Discovery) out of 10	.45 (.73) (0 - 3) CI 0.29 to 0.61	1.78 (1.80) (0 - 6) CI 1.38 to 2.18	1.49 (1.65) (0 - 7) CI 1.12 to 1.86)
No-intervention words out of 10	.15 (.40) (0 - 2) CI 0.06 to 0.24	.08 (.31) (0 - 2) CI 0.01 to 0.15	.14 (.35) (0 - 1) CI 0.06 to 0.22

The level of expressive word use in all conditions was very low and demonstrated floor effects, with all data except post-intervention expressive word use of experimental words being positively skewed. Therefore, non-parametric analyses were used to examine changes in expressive word use over time for each condition. Three separate Friedman's one-way repeated measures ANOVAs were conducted followed by post-hoc Wilcoxon's signed ranks tests with Bonferroni corrections applied. For words taught through *usual teaching practice*, there was a significant effect of time on expressive word use, χ^2 (2) = 7.369, p = .025. Wilcoxon's signed ranks showed that expressive word use of usual teaching practice words was significantly greater at the post-intervention point (M = .96, SD = 1.39) than at pre-intervention (M = .58, SD = .91), Z = 2.674, p = .007, effect size d = 0.33; but that there was no significant change between post-intervention and follow-up (M = .97, SD = 1.37), Z = -.186, p = .853. This indicated that students' expressive word use increased following usual teaching practice, and that this increase was maintained five weeks later. This was confirmed by a significant difference between pre-intervention and follow-up, Z = -3.157, p = .002. For *experimental* words, there was also a significant effect of time on expressive word use, χ^2 (2) = 53.153, p < .001. Wilcoxon's signed ranks showed that expressive word use of experimental words was significantly greater at the post-intervention point (M = 1.78, SD = 1.80) than at pre-intervention (M = .45, SD = .73), Z = -5.783, p < .001, effect size d = 0.96. Expressive word use at follow-up (M = 1.49, SD = 1.65) was significantly lower than at post-intervention, Z = -2.556, p = .011; but still significantly greater than at pre-intervention, Z = -5.398, p < .001. This indicated that students' expressive word use of experimental words increased following the experimental intervention, and that this increase was partially maintained five weeks later. There was no significant effect of time on expressive word use for no-intervention words (χ^2 (2) = 4.192, p = .123).

Three further Friedman's one-way ANOVAs were conducted to examine differences in expressive word use between the three teaching conditions at each time-point, followed by post-hoc Wilcoxon's signed ranks tests, Bonferroni corrected

At *pre-intervention*, there was a significant difference between the teaching conditions in expressive word use (χ^2 (2) = 20.162, p < .001). Wilcoxon's signed ranks showed that there was no difference in expressive word use between the usual teaching practice words (M =

.58, *SD* = .91) and the experimental words (*M* = .45, *SD* = .73), *Z* = -1.059, *p* = .290, but that expressive word use of no-intervention words (*M* = 0.15, *SD* = 0.40) was significantly lower than that of both usual teaching practice words (*Z* = -3.94, *p* < .001) and experimental words (*Z* = -3.41, *p* = .001). There was also a significant difference between the teaching conditions in expressive word use at *post-intervention* (χ^2 (2) = 67.980, p < .001), with expressive word use of experimental words (*M* = 1.78, *SD* = 1.80) being significantly greater than that of usual teaching practice words (*M* = .96, *SD* = 1.39), *Z* = -3.796, *p* < .001. Again, expressive word use of no-intervention words (*M* = 0.08, *SD* = 0.31) was significantly lower than that of both usual teaching practice words (*Z* = -5.35, p < .001) and experimental words (*Z* = -6.33, *p* < .001). At *follow-up*, there was a significant difference between the teaching conditions (χ^2 (2) = 49,922, *p* < .001). Expressive word use of experimental words (*M* = 1.49, *SD* = 1.65) continued to be greater than that of usual teaching practice words (*M* = 0.97, *SD* = 1.37), *Z* = -2.472, *p* = .013. Expressive word use of no-intervention words (*M* = 0.14, *SD* = 0.35) continued to be significantly lower than that of both usual teaching practice words (*Z* = -5.10, p < .001) and experimental words (*Z* = -5.96, *p* < .001).

Discussion

The aim of the current study was to examine the effectiveness of a new vocabulary intervention package, Word Discovery, in increasing the participants' knowledge of science curriculum words. Participants' knowledge of 30 science words was assessed at pre-intervention, post-intervention, and follow-up assessment points: 10 words were taught by science teachers through usual teaching practice; 10 matched experimental words were taught by the same teachers using the experimental intervention activities; and 10 matched words received no intervention.

In respect of hypothesis 1, that the increase in *depth of word knowledge* for experimental words would be greater than that for words taught through usual teaching practice, from pre- to post-intervention, results were in line with predictions, with experimental intervention being more effective than usual teaching practice in increasing the depth of word knowledge of participating students. The mean numerical gain from pre- to post-intervention for depth of word knowledge of usual teaching practice words was 1.58 (*SD* = 2.71; *d* = 0.52, medium effect size), and for experimental words it was 3.46 (*SD* = 3.24; *d* = 1.09, large effect size), out of a possible 20. Regarding maintenance of depth of word knowledge, results were less clearly in line with predictions: from post-intervention to follow-up, depth of word knowledge of usual teaching practice words was maintained, but depth of word knowledge of experimental words was not fully maintained. Thus, hypothesis 1 with regard to maintenance of depth of word knowledge was not supported.

In respect of hypothesis 2, that the increase in *expressive use* of experimental words would be greater than for words taught through usual teaching practice, from pre- to post-intervention, results were again in line with predictions, with experimental intervention being more effective than usual teaching practice in increasing the expressive word use of participating students. The mean numerical gain from pre- to post-intervention for expressive word use of usual teaching practice words was 0.58 (*SD* = 1.21; *d* = 0.33, small effect size), and for experimental words it was 1.33 (*SD* = 1.67; *d* = 0.96, large effect size), out of a possible 10. Regarding maintenance of expressive word use, expressive use of usual teaching practice words was maintained from post-intervention to follow-up, but expressive use of experimental words was not fully maintained. Nonetheless, at follow-up, expressive use of experimental words remained significantly higher than expressive use of usual teaching practice words. Thus, there was partial support for hypothesis 2 with regard to maintenance of expressive word use.

There was no change in students' depth of word knowledge or expressive use of no-intervention words over time, confirming that change in usual teaching practice and experimental word knowledge was not due to maturity or practice effects.

Depth of word knowledge pre- to post-intervention

If we consider the effect of Word Discovery compared with usual teaching practice, the difference in depth of word knowledge gain was 1.88 (out of a possible 20) representing additional knowledge of up to two words in the experimental condition, with the resultant advantages in the classroom. These results are comparable to those found in other vocabulary intervention studies with adolescents. For example, the mean word knowledge gain in Snow et al. (2009) was 4.43 out of 40 assessed words (d = .21, small effect size); and in Spencer et al. (2017), the mean gain was 1.17 out of 10 targeted words (□^{2} = .42, large effect size). The clinical significance of the gains in the current study is further demonstrated by considering that gains were achieved with a smaller amount of cumulative intervention intensity compared to other studies. For example, in the current study, intervention duration ranged from 6.5 minutes to 2.25 hours (mean 62 minutes) over an average of four to five weeks. This compares with cross-curricular intervention throughout the course of one academic year (Snow et al., 2009) and six to seven hours' intervention over 10 weeks (Spencer et al., 2017). Furthermore, as this was a cascading intervention, whereby the SLT trained another agent of change (the teachers), it is relevant to consider the amount of training provided. In the current study, this was one hour, considerably less than in other studies. For example, in a study by Starling et al. (2012), training on language modification techniques took place in 50-minute sessions once a week for 10 weeks, concurrently with the intervention.

Expressive word use pre- to post-intervention

Considering the effect of Word Discovery compared with usual teaching practice, the difference in expressive word use was 0.95, representing an additional expressive advantage of one word (out of a possible 10). While this mirrors the depth of word knowledge result, this result needs to be interpreted with caution. Firstly, the level of expressive word use in all conditions was very low and demonstrated floor effects, and, further, the expressive word use score depended on success on the depth of word knowledge task. It was, however, felt to be important to report the effect of the intervention on expressive word use, given the significance of the ability to use words expressively for examination success, and results do indicate a potential for Word Discovery intervention to impact on expressive word use.

Why was Word Discovery intervention effective?

One reason for the effectiveness of Word Discovery intervention in increasing depth of word knowledge, and potentially expressive word use, may be because it addressed multiple aspects of word learning. In fidelity observations of usual teaching practice during topic 1, limited use of semantic feature analysis or phonological awareness strategies was observed. The survey of mainstream secondary school teachers and speech and language therapists by Lowe et al. (in preparation) suggests that this is typical of vocabulary teaching practice across the UK. In contrast, Word Discovery activities made phonological and semantic information explicit, and linked phonological with semantic information, thus facilitating the mapping of phonological form onto semantic content (Leonard, 1998). In addition, the activities involved deliberate verbal repetition of the words by both teachers and students, thus supporting phonological processing skills. The Word Discovery approach, therefore, had the potential to benefit those who had phonological weaknesses but relative semantic strengths, as well as those for whom the converse was true. In addition, visual support, orthographic input, and personalisation were intrinsic to the intervention, thus exploiting a wide range of modalities and skills.

Furthermore, the experimental words received somewhat more exposure than the usual teaching practice words. Amount of exposure has been shown to be associated with increased word learning for younger children with language disorder (Rice, Oetting, Marquis, Bode, & Pae, 1994). The Word Discovery intervention provided a framework in which word exposure occurred in focused and meaningful contexts, directly targeting the phonological and semantic aspects of word learning which are challenging for children and adolescents with language disorder.

Maintenance

The modest increases in depth of knowledge of the usual teaching practice words were maintained at the follow-up time-point, whereas the larger increases of the experimental Word Discovery words were less well maintained. This was also the case for maintenance of expressive word use, although expressive word use of Word Discovery words remained significantly higher than that of usual teaching practice words. This contrasts with other vocabulary studies, which have demonstrated retention in word knowledge: for example, Clegg (2014) who used a four-week follow-up period; and Spencer et al. (2017), who used a 10-week follow-up period. A possible explanation for the different findings of the current study may relate to verbal working memory and semantic representation weaknesses. Many participants had very low scores on the CELF-4 UK Recalling Sentences subtest and the WMTB-C Listening Recall subtest, implying possible inefficiencies within verbal working memory (Henry & Botting, 2017). A potential hypothesis is that during the word-learning activities, components of working memory were repeatedly activated, for the experimental words more so than for the usual teaching practice words, keeping the experimental words constantly primed, but that due to the verbal working memory and semantic limitations of the participants, less secure or limited semantic representations were laid down. If this was the case, during the follow-up period with no revision, the insecure traces of understanding which had been acquired during the intervention period could have become lost or difficult to retrieve. This interpretation is consistent with the proposition by Kail and Leonard (1986) that semantic limitations contribute to inefficient word storage. A speculative hypothesis to explain why expressive word use showed a greater tendency towards better maintenance than depth of word knowledge, could be that the phonological component of Word Discovery activities facilitated expressive performance by strengthening phonological deficits, and enabling stronger phonological representations to be established. This would concur with the position that weak phonological skills particularly affect naming (Stackhouse & Wells, 1997; Lahey & Edwards, 1999).

However, to test these hypotheses, future work is required, in which more precise information is obtained on the working memory, semantic, and phonological skills of the participants than was sought in the current study.

Clinical and educational implications

The success of the intervention was dependent upon effective collaboration between the researcher (a SLT) and participating science teachers. The use of curriculum vocabulary in the current study conforms to the concept of "curriculum-relevant therapy" advocated by Ehren (2002; p.60), which represents a meeting point at which the different spheres of knowledge of the teacher and the SLT/P can meet and bear fruit. The current study builds on the findings of Murphy et al. (2017), which showed the effectiveness of delivering a programme of vocabulary intervention within English lessons. The added valued provided by the current study is that of applying principles of vocabulary intervention to the science curriculum, targeting words which are inherent to the subject syllabus. This approach has the potential to be applicable to subjects across the whole secondary school curriculum.

The gains in depth of word knowledge and expressive word use were made as a consequence of relatively modest input, both in the amount of training which the teachers received (one hour), and in the amount of intervention which the students received (average one hour). As this amount of teacher / SLT/P collaboration and classroom input was achievable in the research context, it is reasonable to conclude that it has the potential to translate into practice, demonstrating ecological validity.

A further reason for the success of the intervention may have been its timeliness, occurring at a point in time where developmental and environmental opportunities coincide. Developmentally, this relates to continuing development of metalinguistic awareness during adolescence (van Kleeck, 1984; Spencer, Clegg, & Stackhouse, 2013). At the same time, because of the heightened neurological changes taking place during adolescence, students have the potential to respond to educational and rehabilitation programmes during this period (Blakemore & Choudhury, 2006). These developments coincide with a critical period in the adolescent's school career. At secondary school, vocabulary becomes increasingly technical and specialised (Nippold, 2007), yet there is no respite from the pace at which new vocabulary is presented to students, nor from the increasing focus on examination success (DfE, 2017b). Therefore, research and practice in the field of language disorder need to move beyond the view that intervention can be effective even in this older age-group (e.g. Ebbels et al., 2012), to a standpoint where adolescence is viewed as a critical window of opportunity in which it is crucial to intervene.

The purpose of the follow-up assessment was to evaluate retention after a period of no input, and it showed that recently acquired phonological and semantic knowledge had, to some extent, deteriorated. The implication of these findings for practice is the need for constant revision to maintain recently-acquired word knowledge. If revision opportunities occur at intervals that are too far apart, the deterioration of phonological and semantic information could result in the information becoming too poorly specified to retrieve, leaving students with language disorder at risk of falling further behind. Revision opportunities for students with language disorder need to occur with regularity and frequency.

Limitations and areas for future research

One potential confound arising from the study design was that, because the two sets of words were chosen from different topics, some topics may have been more interesting to students than others, and some topics may have had a propensity towards more abstract or technical words than others. However, the fact that there were 22 different sets of words across the study mitigates against these possibilities, which could have arisen if a single set of experimental and control words had been used, adding confidence to the findings.

Blind assessment would have added further to the strength of the study; however, due to the geographical spread of participating schools, and the critical timing for assessments, it was not possible to source and train independent assessors in the given timeframe. In the absence of blind assessment, independent reliability checks were employed to maximise the rigour of the results, with the resultant kappa coefficients indicating strong inter-rater reliability.

The presence of floor effects in the expressive word use measure was possibly a consequence of the complexity of the subject-specific vocabulary. So that indicative results could be obtained, this was dealt with by the use of non-parametric statistics. Nonetheless, replication of the study is necessary using word sets which are more within participants' zone of proximal development.

Other areas for future research include examination of the impact of working memory, semantic, and phonological skills of the participants on their response to intervention, as well as the differential effects of each component word-learning activity. It is also important, given the prevalence of low language levels in areas of social disadvantage, to investigate the impact of socio-economic status on the results of intervention.

Conclusion

The findings of the current study underline the considerable difficulty which adolescents with language disorder have in understanding and using science curriculum words. This lends strong support to the findings of previous research showing the persistence of language disorder in adolescence, and the complexity of science vocabulary, particularly for adolescents with language disorder. Because of these ongoing difficulties, it is essential to find optimum ways of helping these students. The Word Discovery intervention in the current study led to positive gains in depth of word knowledge and expressive word use immediately following intervention, which were maintained to a more limited degree at follow-up. It, therefore, represents a candidate for inclusion in the intervention options open to teachers and speech and language therapists/pathologists working collaboratively in secondary schools.

Acknowledgements

The authors thank the participating staff and students from all the schools involved.

References

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423.

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

Blakemore, S., & Choudhury, S. (2006). Development of the adolescent brain: implications for executive function and social cognition. *Journal of Child Psychology & Psychiatry*, *37*(3-4), 296–312.

Bishop, D., Snowling, M., Thompson, P., Greenhalgh, T., & CATALISE-2 consortium. (2017). Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology & Psychiatry*.

Bishop, D. V. M. (2014). *Uncommon Understanding*. London: Psychology Press.

Brownell, R. (2010). Expressive One-Word Picture Vocabulary Test. San Antonio, TX: Pearson.

Clegg, J. (2014). Curriculum vocabulary learning: intervention for children with social, emotional and behavioural difficulties (SEBD): findings from a case study series. *Emotional and Behavioural Difficulties*, *19*(1), 106–127.

Crumpler, M., & McCarty, C. (2006). Access Reading Test. Hodder Education.

Dale, E. (1965). Vocabulary Measurement: Techniques and Major Findings. *Elementary English*, *42*(8), 895–948.

Department for Communities and Local Government. (2015). The English Index of Multiple Deprivation. London UK. Retrieved April 2, 2017, from http://dclgapps.communities.gov.uk/imd/idmap.html

Department for Education (2014). The National Curriculum in England: Key stages 3 and 4 framework document. London UK. https://doi.org/https://www.gov.uk/government/collections/national-curriculum

Department for Education (2016). *Guidance: pupil premium 2016 – 2017: conditions of grant*. London UK.

Department for Education (2017a). EduBase2. Retrieved January 26, 2017, from http://www.education.gov.uk/edubase/establishment/establishmentdetails

Department for Education (2017b). *Progress 8 and Attainment 8 Guide for maintained secondary schools, academies and free schools*. London UK.

Department for Education (2018) *Schools, pupils and their characteristics: January 2018* London UK.

Dockrell, J. E., Braisby, N., & Best, R. M. (2007). Children's acquisition of science terms: Simple exposure is insufficient. *Learning and Instruction*, *17*(6), 577–594.

Dunn, L., Dunn, L., Sewell, J., & Styles, B. (2011). British Picture Vocabulary Scale, third edition. Granada.

Ehren, B. (2002). Speech-Language Pathologists Contributing Significantly to the Academic Success of High School Students: A Vision for Professional Growth. *Topics in Language Disorders*, *22*(2), 60–80.

Ebbels, S. H., Nicoll, H., Clark, B., Eachus, B., Gallagher, A. L., Horniman, K., … Turner, G. (2012). Effectiveness of semantic therapy for word-finding difficulties in pupils with persistent language impairments: a randomized control trial. *International Journal of Language & Communication Disorders*, *47*(1), 35–51.

Elks, L., & McLachlan, H. (2008). *Secondary Language Builders: Speech and Language Support for 11 – 16s*. St Mabyn: Elklan.

Ford-Connors, E., & Paratore, J. (2015). Vocabulary Instruction in Fifth Grade and Beyond: Sources of Word Learning and Productive Contexts for Development. *Review of Educational Research*, *85*(1), 50–91.

Forwood, C. (2014). Beaker or flask: what's in a word? In *Poster presentation: Mind the Gap: Putting research into practice RCSLT Conference*.

Frederickson, N., Frith, U., & Reason, R. (1997). Phonological Awareness Battery. Chesterfield: Winslow.

Gathercole, S., & Pickering, S. (2001). Working Memory Test Battery for Children. Pearson.

German, D. (1990). Test of Adolescent/Adult Word Finding . Chicago: Riverside.

German, D. (1991). Test of Word Finding in Discourse. DLM Resources.

GL Assessment. (2015). Cognitive Attainment Test (CAT). GL Assessment.

Henry, L., & Botting, N. (2017). Working memory and developmental language impairments. *Child Language Teaching and Therapy*, *33*(1), 19–32.

Hollands, K., van Kraayenoord, C., & McMahon, S. (2005). Support to adolescents experiencing language difficulties: A survey of speech-language pathologists. *Advances in Speech–Language Pathology*, *7*(3), 113–129.

Joffe, V. L. (2011). *Vocabulary Enrichment Programme*. Milton Keynes: Speechmark Publishers.

Johnson, C. J., Beitchman, J. H., & Brownlie, E. B. (2010). Twenty-Year Follow-Up of Children With and Without Speech-Language Impairments: Family, Educational, Occupational, and Quality of Life Outcomes. *American Journal of Speech Language Pathology*, *19*, 51–65.

Kail, R., & Leonard, L. (1986). Word-finding abilities in language-impaired children . *ASHA Monographs*, *25*.

Lahey, M., & Edwards, J. (1999). Naming Errors of Children with Specific Language Impairment. *Journal of Speech, Language, and Hearing Research*, *42*, 195–205.

Leonard, L. (1998). *Children with Specific Language Impairment*. Cambridge, MA: The MIT Press.

Lindsay, G., Dockrell, J., Mackie, C., & Letchford, B. (2005). Local education authorities' approaches to provision for children with specific speech and language difficulties in England and Wales. *European Journal of Special Needs Education*, *20*(3), 329–345.

Lowe, H., Henry, L., Müller, L.-M., & Joffe, V. L. (2017). Vocabulary intervention for adolescents with language disorder: a systematic review. *International Journal of Language & Communication Disorders*.

Lowe, H., Henry, L., Wallinger, J., & Joffe, V. L. (in preparation). Teaching vocabulary to adolescents with language disorder in mainstream secondary schools: perspectives from teachers and speech and language therapists.

Lowe, H., & Joffe, V. (2017). Exploring the feasibility of a classroom-based vocabulary intervention for mainstream secondary school students with language disorder. *Support for Learning*, *32*(2), 110–128.

Lowry, R. (2001-2017). Vassar Stats. Retrieved June 18, 2017, from http://vassarstats.net/kappa.html

MacSwan, J., & Pray, L. (2005). Learning English Bilingually: Age of Onset of Exposure and Rate of Acquisition Among English Language Learners in a Bilingual Education Program. *Bilingual Research Journal*, 29(3), 653.

McArthur, G. M., Hogben, J. H., Edwards, V. T., Heath, S. M., & Mengler, E. D. (2000). On the ``specifics’’ of specific reading disability and specific language impairment. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *41*(7), 869–874.

McGregor, K. K., Newman, R. M., Reilly, R. M., & Capone, N. C. (2002). Semantic representation and naming in children with specific language impairment. *Journal of Speech, Language, and Hearing Research,* 45(5), 998–1015.

McGregor K., K., Oleson, J., Bahnsen, A., & Duff, D. (2013). Children with developmental language impairment have vocabulary deficits characterized by limited breadth and depth. *International Journal of Language & Communication Disorders*, *48*(3), 307–319.

Murphy, A., Franklin, S., Breen, A., Hanlon, M., McNamara, A., Bogue, A., & James, E. (2017). A whole class teaching approach to improve the vocabulary skills of adolescents attending mainstream secondary school, in areas of socioeconomic disadvantage. *Child Language Teaching and Therapy*, *33*(2), 129–144.

Nash, H., & Snowling, M. (2006). Teaching new words to children with poor existing vocabulary knowledge: a controlled evaluation of the definition and context methods. *International Journal of Language & Communication Disorders*, *41*(3), 335–354.

Nation, K., & Snowling, M. J. (2004). Beyond phonological skills: broader language skills contribute to the development of reading. *Journal of Research in Reading*, *27*(4), 342–356.

Nippold, M. A. (2007). *Later language development: School-age children, adolescents, and young adults* (3rd ed.). Austin, TX: Pro-Ed.

Ofsted. (2017). Ofsted Inspection Reports. Retrieved January 25, 2017, from https://reports.ofsted.gov.uk/inspection-reports/find-inspection-report

Rice, M. L., Oetting, J. B., Marquis, J., Bode, J., & Pae, S. (1994). Frequency of input effects on word comprehension of children with specific language impairment. *Journal of Speech and Hearing Research*, *37*(1), 106–122.

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.

Royal College of Speech and Language Therapists/ICAN. (2018). *Bercow: Ten Years On*.

Semel, E., Wiig, E., & Secord, W. A. (2006). Clinical evaluation of language fundamentals UK 4th edition. London: Pearson.

Snow, C. E., Lawrence, J. F., & White, C. E. (2009). Generating Knowledge of Academic Language Among Urban Middle School Students. *Journal of Research on Educational Effectiveness*, *2*(4), 4,325–344.

Spencer, S., Clegg, J., & Stackhouse, J. (2013). Language, social class and education: listening to adolescents' perceptions. *Language and Education*, *27*(2), 129–143.

Spencer, S., Clegg, J., Lowe, H., & Stackhouse, J. (2017). Increasing adolescents' depth of understanding of cross-curriculum words: an intervention study. International Journal of Language and Communication Disorders, 52(5) 652-668

Stackhouse, J., & Wells, B. (1997). *Children's Speech and Literacy Impairment*. London: Whurr.

Starling, J., Munro, N., Togher, L., & Arciuli, J. (2012). Training Secondary School Teachers in Instructional Language Modification Techniques to Support Adolescents With Language Impairment: A Randomized Controlled Trial. *Language, Speech and Hearing Services in Schools*, *43*(4), 474–495.

Throneburg, R. N., Calvert, L. K., Sturm, J. J., Paramboukas, A. a., & Paul, P. J. (2000). A Comparison of Service Delivery Models: Effects on Curricular Vocabulary Skills in the School Setting. *American Journal of Speech-Language Pathology*, *9*(1), 10–20.

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X. Y., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech Language and Hearing Research*, *40*(6), 1245–1260.

Van Heuven, W., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190.

Van Kleeck, A. (1984). Metalinguistic skills: cutting across spoken and written language and problem-solving abilities. In G. P. Wallach & K. Butler (Eds.), *Language Learning Disabilities in School-age Children* (pp. 128–153). Baltimore/London: Williams and Wilkins.

Warren, S., Fey, M., & Yoder, P. (2007). Differential Treatment Intensity Research: A Missing Link to Creating Optimally Effective Communication Interventions. *Mental Retardation and Developmental Disabilities Research Reviews*, *13*, 70–77.

Wechsler, D. (2011). Wechsler Abbreviated Scale of Intelligence, second edition. USA: Pearson Clinical.

Woodward, J., & Noell, J. (1991). Science instruction at the secondary level: implications for students with learning disabilities. *Journal of Learning Disability*, *24*(5), 277–284.