International Journal of Language & Communication Disorders

The effectiveness of a small group vocabulary intervention programme: evidence from a Regression Discontinuity Design

Journal:	*International Journal of Language & Communication Disorders*
Manuscript ID	TLCD-2016-0179.R3
Wiley - Manuscript type:	Research Report
Keywords:	vocabulary, intervention, school-age children, education, language development

EFFECTIVENESS OF A VOCABULARY PROGRAMME

Abstract

Purpose: Children's vocabulary knowledge is closely related to other measures of language development and to literacy skills and educational attainment. This study uses a Regression Discontinuity Design (RDD) to evaluate the effectiveness of a small group vocabulary intervention programme for children with poor vocabulary knowledge.

Method: The vocabulary knowledge of children aged 6 to 9 years was assessed in six classes. Based on scores at initial assessment, children with low vocabulary scores for their age were assigned to an intervention group (43 children), with the remaining 156 children assigned to a control group. Children in the intervention group received ~~a~~ 2 – 3 small group weekly teaching sessions over a 10 week period. ~~10-week intervention programme delivered in small groups~~. All children were re-tested post-intervention on the same measures of vocabulary knowledge.

Results: The intervention group showed significant improvements in their knowledge of the meanings of the taught words at post-test (an additional 3.95 words learned [95% CI 2.70 – 5.20] compared to control group; *d* = 1.20), but the effects did not generalize to untaught words.

Conclusions: A small group vocabulary intervention programme is effective for teaching word meanings to 6- to 9-year-old children with poor vocabulary skills. This study provides further evidence that the regression discontinuity design is an effective method for evaluating educational interventions.

EFFECTIVENESS OF A VOCABULARY PROGRAMME

What is already known on this subject: Children's vocabulary knowledge is closely related to other measures of language development and to literacy skills and educational attainment. A variety of vocabulary intervention programmes have previously been evaluated and teaching words explicitly (direct instruction) has been found to be effective. The Regression Discontinuity Design has been proposed as an effective way to evaluate educational interventions.

What this study adds: This study uses a Regression Discontinuity Design to evaluate a programme which uses in-depth, direct instruction to teach children the meanings of words. The results show that the vocabulary intervention programme was effective in improving understanding of the directly taught words, but did not lead to generalized improvements to untaught words. The study shows that the Regression Discontinuity Design is useful for evaluating the effectiveness of educational interventions because it provides evidence for the effectiveness of an intervention whilst also allowing every child who requires intervention to receive it.

Clinical implications of this study: This study demonstrates that direct, in-depth instruction can be effective for teaching the meanings of words to children with poor vocabulary knowledge, although as might be expected, this approach does not produce generalized improvements to untaught vocabulary. Nevertheless, we believe that this approach can be recommended to speech and language therapists and education professionals as a way of teaching vocabulary.

The effectiveness of a small group vocabulary intervention programme: Evidence from a Regression Discontinuity Design

Words are the building blocks of language. A typical adult is estimated to know between 45,000 and 80,000 words (Nagy & Herman, 1987) which equates to learning approximately 10 words every day between the ages of 1 to 18 years (Bloom, 2000).

Children learn words naturally from their linguistic environment, irrespective of whether explicit instruction is provided (Beck, McKeown, & Kucan, 2002). However, relying solely on incidental learning as the basis for children's vocabulary development is problematic, because children experience linguistic environments that differ widely in quality, and rates of learning differ markedly between children. These differences in environmental input and learning abilities lead to large variations in vocabulary knowledge amongst children that emerge in the preschool years. Poor vocabulary knowledge, in turn, adversely affects language, literacy and educational outcomes.

The relationship between vocabulary levels and educational attainment

Literacy levels, and educational attainments more broadly, are highly correlated with children's vocabulary knowledge. In a longitudinal study, Lee (2011) found that expressive vocabulary size at age 2 years predicted language and literacy outcomes at age 11 years, even after controlling for gender, ethnicity and SES. Similarly, Marchman and Fernald (2008) found that speed of spoken word recognition and vocabulary size at 24 months predicted linguistic and cognitive skills (spatial reasoning, pattern matching and working memory) at 8 years. Vocabulary also predicts educational outcomes; in a large-scale assessment of school readiness in Canada, children's vocabulary levels when starting school were one of the

strongest predictors of later educational achievement (Romano, Babchishin, Pagani, & Kohen, 2010).

Vocabulary knowledge is a strong predictor of children's literacy development. Early vocabulary knowledge predicts later reading comprehension abilities (Lervag, Hulme & Melby-Lervag, 2017; Muter et al., 2004) and children at risk of reading comprehension problems show slower rates of vocabulary development than controls (Cain & Oakhill, 2011). Furthermore, oral language interventions can improve reading comprehension skills in children with reading comprehension difficulties; an effect that is partially mediated by improvements in vocabulary knowledge (Clarke et al., 2010). Finally, a broad measure of oral language (which includes measures of vocabulary knowledge) at 3½ years predicts early word-level reading in the first year of school as well as variations in reading comprehension at 8–9 years of age (Hulme et al., 2015). The fact that children from disadvantaged backgrounds show poorer vocabulary knowledge, and that such vocabulary weaknesses appear to be causally related to the development of reading skills provides a strong justification for advocating the use of vocabulary instruction in educational settings (Beck et al., 2002; Biemiller, 2003).

Vocabulary instruction

There are broadly two approaches to vocabulary instruction (implicit vs. explicit) although in practice many research studies use a combination of these approaches. In implicit instruction, exposure to new words is embedded within an activity, such as reading a story. Explicit instruction, in contrast, involves teaching the meanings of words directly, by providing definitions.

Biemiller and Boote (2006) provide evidence for the superiority of explaining definitions over implicit exposure to new words. They showed that children acquired

an understanding of 12% of key words from a story during repetitive readings, but that an extra 10% of words were learned when explanations of word meanings were added to the story reading. Similarly, Dockrell, Braisby and Best (2007) showed that children who were simply exposed to new science terms without any instruction had a low rate of retention and understanding of the meanings of those words. This is echoed by Kirschner, Sweller, and Clarke (2006) who argue that minimal guidance during instruction is inferior to guided instruction in terms of the amount and accuracy of information learned.

Coyne, McCoach, Loftus, Zipoli, and Kapp (2009) directly compared embedded, implicit instruction and explicit teaching of words through storybook reading in preschool. They demonstrated that children who were taught explicitly demonstrated deeper, more refined vocabulary knowledge than those in the implicit vocabulary learning condition. For children with developmental language disorder, Storkel et al., 2016 reported a vocabulary training study based around interactive book reading and found that a relatively large number of exposures to new words was needed to bring about learning. Steele, Willoughby and Mills (2013) conducted a small-scale experimental vocabulary training study, children with developmental language disorder learned the meanings of new words less well than age-matched controls but benefitted when give explicit instruction in the semantic properties of new words. Direct instruction therefore appears to improve vocabulary learning in both typically developing children and those with developmental language disorder. In two reviews, Marulis and Neuman (2010) and Christ and Wang (2011) argue that combining direct instruction with interactive activities, such as encouraging children to generate their own examples of word use, is the most effective method for teaching new words.

Current knowledge therefore suggest that the best way to ensure the learning and retention of the meanings of new words is to combine explicit teaching of definitions with interactive activities which encourage word use in a variety of contexts. When advocating direct instruction, Beck et al., (2002) classify words into three tiers to help with selection of words to teach. They suggest tier two words, which children are likely to encounter in both oral language and reading, and which provide more complex synonyms for familiar concepts (e.g. *anxious* for *scared* or *worried*) should be the focus of classroom instructional programmes.

The intervention materials used for this study were designed using the most effective aspects of the intervention programmes reviewed by Christ and Wang (2011) and Marulis and Neuman (2010), combined with the teaching principles advocated by Beck et al. (2002). This resulted in a teaching programme involving in-depth, direct instruction on a small number of tier two target words using a variety of teaching methods, including explicit teaching of definitions and practicing the use of target words in multiple contexts. The materials developed were tested for their effectiveness in teaching the meanings of new vocabulary items to 6- to 9-year-old children with relatively poor vocabulary skills compared to their peers.

The Regression Discontinuity Design (RDD)

To evaluate the effectiveness of our vocabulary teaching programme we used a Regression Discontinuity Design (RDD). The RDD is a quasi-experimental design that analyses pre-intervention against post-intervention scores in the context of a known cut-off point for intervention group assignment. A treatment effect is shown when there is a discontinuity in the regression line at the cut-off point for group assignment (Cappelleri & Trochim, 2003). If the regression function for the intervention group is higher than that for the control group this provides evidence for

a causal effect of the intervention in increasing post-test scores. In practical terms, the RDD involves comparing children with lower attainment levels on a specific measure to their more able peers before and after an intervention programme has been given to the lower attaining group only.

The logic of an RDD is best understood graphically. Figure 1a shows simulated data for children's language scores at two times of testing without any intervention. The correlation between language scores at the two times is r= .6. The graph shows the best fitting regression line for all points; there is no discontinuity in the regression function. Figure 1b shows data assuming that children with a pretest score of less than 0, have received a language intervention which, on average has increased their language scores by 2 points. There is a clear discontinuity in this graph between the function for the intervention group and the function for the control group: posttest scores for the intervention group are higher than expected given their pretest scores. The difference in the height of the regression functions for the two groups provides an estimate of the size of the intervention effect (2 points in this case).

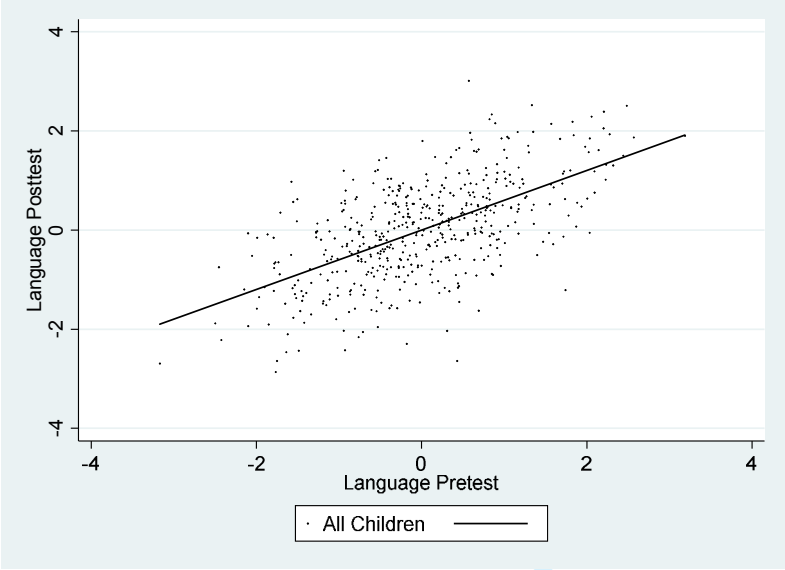


Figure 1a: Simulated data for 500 children showing the hypothetical regression function relating pretest to posttest language scores

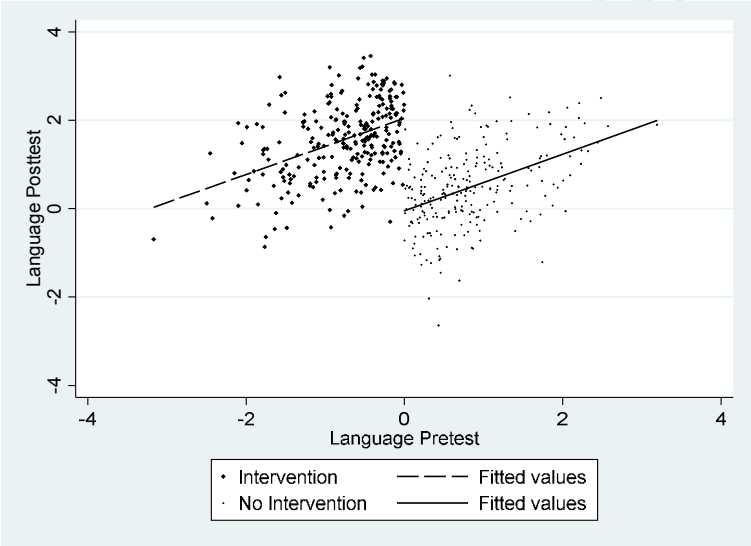


Figure 1b: Simulated data for 500 children showing the hypothetical regression function relating pretest to posttest language scores, where children with a score below zero have received effective intervention

Many have advocated the benefits of the RDD in educational settings (e.g. Ludwig & Miller, 2007). One advantage of this design is that participants are assigned to an intervention group on the basis of a pre-intervention score; a cut-off point is chosen and all children below that specified level receive the intervention. This means that, in contrast to a Randomized Controlled Trial, all children who require help receive it (Trochim, 2006). However, one potential disadvantage of the RDD is that statistical power to detect treatment effects is lower than in a Randomised Control Trial (RCT; Goldberger, 1972). In addition, the analysis of data from a Regression Discontinuity Design requires accurate specification of the regression model in order to avoid under- or over-estimating the effects of treatment (Trochim, 2006).

The current study tests the efficacy of a programme for teaching vocabulary knowledge. Our target population are children aged 6 to 9 years old with relatively weak vocabulary scores for their age. We expect the programme to produce improvements in knowledge of directly taught words but not to generalize to untaught words. The aims of this study were twofold: to evaluate the effectiveness of a small group vocabulary teaching programme designed to be suitable for low-performing students in mainstream schools and to provide additional evidence concerning the usefulness of the RDD as a way of evaluating educational interventions.

Method

Children were allocated to an intervention or control group based on their age adjusted pre-intervention vocabulary score (see below for further details). Those allocated to the intervention group received 10 weeks of in-depth vocabulary instruction. The same vocabulary measures were used to test both groups of

children pre- and post-intervention. Ethical approval was obtained from the Ethics Committee at University College London, and informed consent was obtained from the head-teachers of the schools to use the vocabulary programme as part of the school curriculum. The data reported here come from two researchers who used identical methods - the first author and a graduate student who was trained to deliver the vocabulary intervention programme and in the assessment methods used. The first author observed the graduate student delivering teaching sessions to ensure that the same protocol was followed. The programme was manualized, further information about this vocabulary programme, including session plans and resources, can be requested by emailing wonderfulwordsintervention@gmail.com.

Participants, Group Allocation and Power

Children (N =199) were recruited from three schools; an independent school and two state-funded schools. Children in years 2 to 4 (aged 6 to 9 years) took part in the study since this was the age range the materials were designed for. All children in each participating classroom were tested using group-administered vocabulary measures before the intervention began (pretest). Each child's total score on the two experimental words sets combined was regressed on age. Children whose unstandardized residuals were less than -4 were allocated to the intervention group. This cut-off represented roughly the bottom quartile of the distribution and gave 43 intervention and 156 control children. The mean ages for the intervention and control group were 7 years 8 months and 7 years 7 months respectively. After a 10-week intervention programme which was given to the intervention group only, all children were retested using the same measures (time 2).

Power in an RDD is lower than in an RCT. Goldberger (1972) calculated that the relative efficiency of an RCT compared to an RDD is 2.75 (if the assignment variable is normally distributed as it is in this case). This means that the sample size for an RDD needs to be 2.75 times the size in an RCT to detect the same standardized mean difference (SMD). The power in an RDD (as in an RCT using ANCOVA to analyze it) depends in part on the proportion of variance in the outcome measure accounted for by the covariates (in this case the pretest or baseline score only). In our case the correlation between pretest and posttest scores is high (R2 = .53 for the untreated control words). Based on calculations presented by Lee and Munk (2008) the current study would need a sample size of 148 to have greater than 80% power (p = .05) to detect an SMD of .5 given that 50% of the variance in posttest scores are accounted by pretest scores. In short, the current study has adequate power to detect medium-sized effects.

Measures

Assessment. Children were assessed using three group-administered measures. For each measure, children were given individual answer sheets to complete. For each test word, four pictures were projected onto a screen (three incorrect items and one portraying the target word). The tester read aloud the target word and children marked their chosen answer on an individual answer paper. For abstract items where pictorial representation was not possible, written definitions were used instead of pictures. In these cases, children had to choose from one of four possible written definitions to match the target word. When written definitions were used, the tester read aloud the target word and subsequently read aloud each of the definitions to ensure children with poorer reading skills were not disadvantaged. It should be noted that pre-tests and post-tests were administered

by the same people who had delivered the intervention; arguably the use of group testing minimizes, but does not totally eliminate, the possibility of bias in the assessments.

The following measures were used:

1. A shortened group administered version of the British Picture Vocabulary Scales (BPVS) using pictures to represent the target words was developed. Sets of 4 pictures were projected onto a screen at the front of the class and the tester said a word; children circled the number of the picture that matched the word in a response booklet. There were 3 practice items, and 33 test items that spanned a wide range of difficulty levels. No norms are available for this shortened test. ~~The words began at a simple level for the age group tested and gradually increased in difficulty to ensure an adequate baseline and ceiling for all children.~~

2. Taught words: the words which were to be taught during the intervention programme (see Table 1). Target words were represented by pictures where possible, but written definitions were also used in the case of abstract words. A sample of the test is given in appendix 1.

3. Untaught words: words matched in difficulty to the taught word set but which were not taught in the intervention. As with the taught words, a combination of pictures and written definitions were used.

Vocabulary Teaching Programme

A vocabulary programme (*Wonderful Words)* was designed using the principles outlined by Beck et al. (2002), both in the choice of words to teach and the instructional methods used. ~~Further information about this programme, including session plans and resources, can be requested by emailing wonderfulwordsintervention@gmail.com.~~

Choice of words. The words taught were chosen using the guidelines provided by Beck et al. (2002). These authors use a three-tier framework to identify target words. Tier 1 words are described as basic vocabulary that children are likely to encounter on a daily basis in oral language. Tier 2 words are those which are more advanced and often found in written language; crucially, they also provide advanced labels for concepts with which children will already be familiar (e.g. anxious for scared). Tier 3 words consist of vocabulary used for specific topics, e.g. scientific vocabulary. Beck et al. (2002) recommend that tier 2 words should be taught in vocabulary intervention programmes since these words are likely to be useful to children in a variety of contexts.

Using this framework, Tier 2 words were selected from a database containing age of acquisition and imageability ratings for an extensive range of words (Bird, Franklin and Howard, 2001). The words selected were rated as being acquired 1-2 years later than the age-group (children up to 9 years old) targeted by the intervention. This meant that the children were unlikely to know the words being taught, but that they would be likely to encounter them later in their educational development. The two word sets each consisted of 20 words which were matched as far as possible on word class, age of acquisition and imageability. The words chosen are shown in Table 1.

Table 1: Taught and untaught word sets

Taught Words		Untaught Words	
fragile	depart	pleasant	assist
gratitude	despise	discomfort	insult

Formatted: Left

vague	anticipate	eventually	mislead
frustration	scarce	recover	obviously
priority	absolutely	fascinating	integrity
envy	subtle	implication	mature
helpless	contemplate	humanity	introduce
resist	vigorous	participate	reliable
survive	clarity	hideous	generous
immediately	anxious	vicious	flourish

Teaching methods.

The teaching methods used focused on explicit, rich instruction (Beck & McKeown 2007; Dockrell et al., 2007). The principles that shaped the activities used are as follows:

Each session began with teacher-led activities to ensure children gained a correct and secure understanding of the target word before engaging with the word as a group and individually

"Student-friendly" definitions were used, as opposed to dictionary definitions, which often give children little help in understanding a word's meaning

The use of multiple examples by both teacher and child to encompass the full range of meanings and contexts in which the target word can be used

The use of multiple modes of input (visual, auditory, graphemic) to ensure maximal retention of the target word and its meaning.

Appendix 2 provides an example of these principles applied to the word fragile.

Intervention format. The children receiving the intervention took part in group teaching sessions containing up to 8 children. Sessions were conducted by

the researchers (1 researcher working in 1 school, another working in 2 schools), and took place two or three times per week over ten weeks for 20-25 minutes per session. *Children received on average 26 sessions (range 20 to 30; standard deviation 4.89)*

Each session provided deep instruction on a single target word in the format outlined below.

The target word was introduced in written and spoken form. The children were encouraged to repeat the word aloud. The definition of the word was read aloud by the researcher. The word was then presented in multiple contexts using the paragraphs and sentences provided in the session plans, with an emphasis on the target word. The children were encouraged to identify when they heard the target word.

Word games were used to allow children to identify and use the target word in different contexts. The first activity in this section usually involved identifying which situations were positively associated with the target word from examples read by the researcher (e.g. which of these objects is fragile?) The second activity provided an extension of the first, whereby the children generated their own examples using the target word. Examples that were incorrect or vague were corrected or clarified by the researcher as appropriate.

In addition, children were given word map worksheets to complete. These consisted of writing the target word, identifying a picture, choosing synonyms and antonyms to match the target word, and an optional extension activity consisting of writing their own examples. While the children completed the worksheet, the researcher was available to answer any questions they had and to talk to them about why they had chosen the answer given. Children who completed the worksheet earlier than others were asked to draw a picture or write a sentence using the target

word. Once all children had completed the worksheet, stickers were handed out and the children were taken back to class.

Results

Table 2: Means, ~~and~~ standard deviations (SD) and maximum scores for intervention and control group pre- and post-intervention

	Intervention Group				Control Group			
	Pre-test		Post-test		Pre-test		Post-test	
~~Variable (Maximum Score)~~	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Taught words (20)*	6.93	2.00	13.28	2.70	11.33	2.30	12.14	2.55
Untaught words (20)*	7.70	3.11	9.53	2.94	13.00	2.56	13.79	2.56
BPVS (33)*	22.03	4.07	24.35	3.76	25.57	3.68	27.25	2.74

*Maximum possible score

Formatted: Indent: First line: 0"

Table 2 shows the raw scores for each group before and after the intervention on the taught and untaught words. The key question is whether children in the intervention group achieved higher scores on the post-intervention measure of the taught words than would be expected from their pre-intervention scores. Evidence for this comes from a discontinuity in the regression function at the point specifying group membership when the pre-test and post-test scores are plotted against each other for both groups. In the analysis of the results, the predictor was centered at the discontinuity point so that the intercept for the control group falls at zero on the x-axis. In practical terms, this means that the cut-off score was adjusted on the graphs, so that all subjects in the intervention group are displayed using negative numbers (less than 0), whilst those in the control group are displayed using positive numbers (greater than 0).

EFFECTIVENESS OF A VOCABULARY PROGRAMME

Improvement on Treated Words

Data for the treated words are plotted in Figure 2. This figure shows children's pre-intervention age-residualised vocabulary scores plotted against their post-intervention scores for the identification of the taught words.

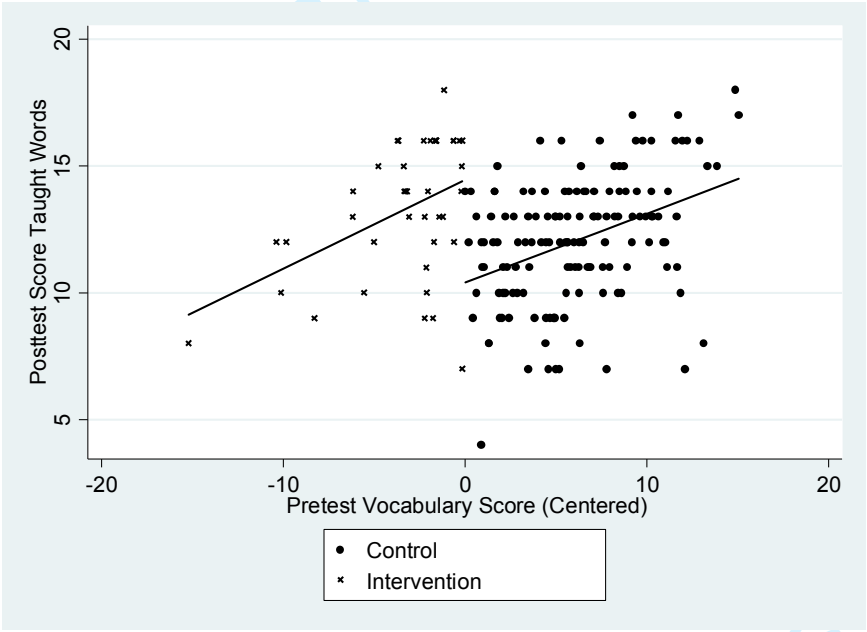


Figure 2: Centred retest vocabulary scores plotted against post-intervention scores by group for treated words (word set 1)

There is a clear discontinuity between the regression functions for the treated children and untreated children. Furthermore, the regression functions appear to be both linear and parallel to each other. To analyze these data, group was coded using a dummy variable (0 = no intervention; 1 = intervention). The pre-test scores and pre-test scores squared were added as predictors to the model to assess whether the function relating pre-test to post-test scores was linear or involved a quadratic components. In addition, to test for interactions between group and the linear and quadratic components, group*pre-test and group*pre-test squared interaction terms were added to the model, in accordance with recommendations

made by Trochim (2006). Each of the non-significant terms was dropped in turn in order to arrive at the model that provided the most parsimonious fit to the data. The final retained model simply included group and the pre-test scores as predictors, and revealed a significant difference between the intercepts of the groups (difference = 3.95 points [95% CI 2.70 – 5.20]; t = 6.23; p = .001). This confirms that the programme was effective in teaching the target words to the children in the intervention group and showed a large effect size (*d* = 1.20; calculated as the difference in intercepts between groups divided by the pooled standard deviation for the two groups at pretest).

Generalization

The same analyses were performed on the scores for the untreated word sets (the matched untaught word-set and the BPVS). The same data modelling approach as for the analysis of data for the treated words was followed. Once again the final simplified models contained only pre-test scores and group as predictors, as analyses showed that inclusion of the quadratic term and the interactions between the group and linear or quadratic terms were not necessary. For the BPVS there was no sign of a group difference in intercept (group difference = .086 [95% CI -1.41 – 1.59]; t = 0.11; p = .91; *d* = 0.02), nor was there any sign of a difference in intercept for the untaught word set (group difference = -0.74; [95% CI -1.19 – 0.48]; t = - 1.19; p = .23; *d* = -0.22). The lack of any significant discontinuity for both these variables are shown in Figures 2 and 3. In summary, the training effects obtained for the taught words did not genrealize to words not directly taught.

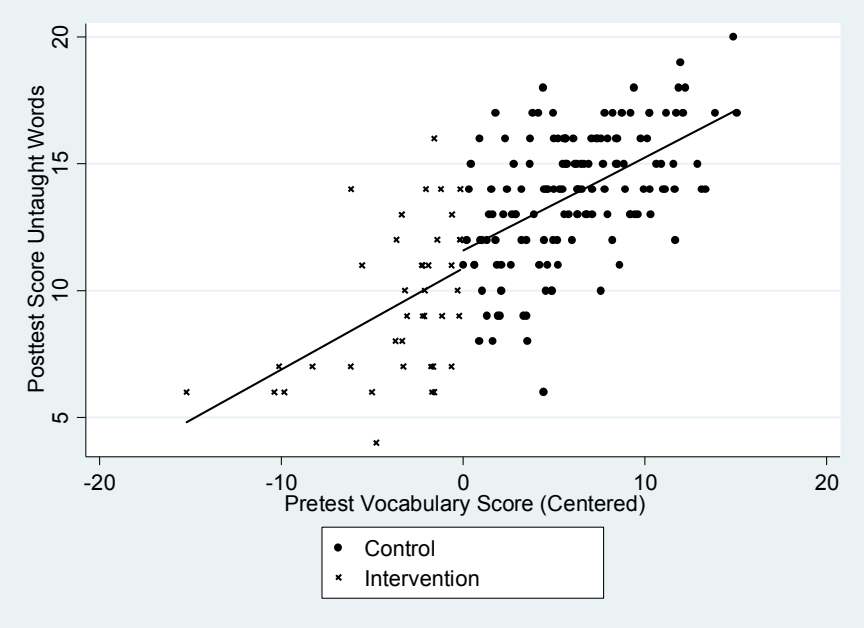


Figure 3: Centred pretest vocabulary scores plotted against post-intervention scores by group for the untaught control words

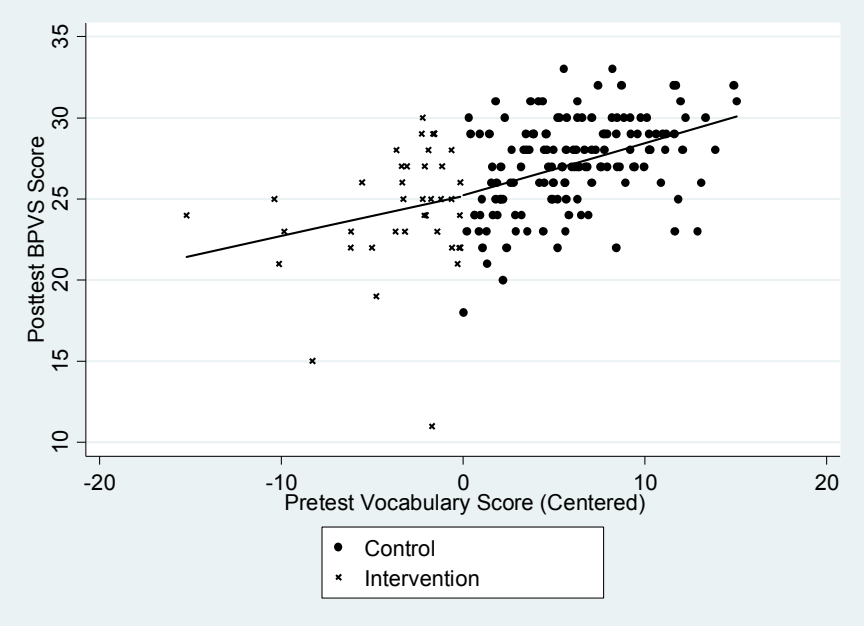


Figure 4: Centred pretest vocabulary scores plotted against post-intervention scores by group for the BPVS

Discussion

Our results demonstrate that ~~our~~ the vocabulary programme was effective in increasing knowledge of directly taught words, to a group of children with relatively weak vocabulary skills in mainstream school settings. The sample in this study was unselected for ability, and the approach used should be effective in typical mainstream school settings. We speculate that the combination of giving children direct instruction using "user-friendly" definitions of words coupled with the multi-sensory learning approach adopted here underlies the effectiveness of the programme. In contrast, there was no sign of generalization to matched untaught words, or to children's knowledge of items taken from the British Picture Vocabulary Scale. On a methodological note, our study shows that the Regression Discontinuity Design was effective in detecting increases in vocabulary knowledge for the words taught in the programme. This provides further evidence that this design is a useful way for evaluating other intervention programmes, perhaps particularly, as in this case, for programmes that can be delivered to groups within school settings.

One question that arises from this study concerns the ratio of teacher input to children's gains on the taught words. This is highlighted by Christ and Wang (2011) who state that teachers must be realistic about the aims of their intervention given the amount of time invested in teaching. The figures in Table 1 show that the children in the intervention group learned approximately 5 new words from the taught set. These gains are in line with those shown by other studies examined in reviews of vocabulary teaching programmes (Christ & Wang, 2011; Marulis & Neuman,

2010). In practical terms, however, this is arguably a small gain from 10 weeks of small group teaching.

Testing

The measure of vocabulary learning used here was a forced choice receptive vocabulary measure. However, as Pearson, Hiebert, and Kamil (2007) argue, asking children to choose a picture to match a target word is a fairly blunt way of measuring vocabulary knowledge. The spectrum of "correct" answers in this type of test ranges from a guess (with a one in four chance of obtaining a correct answer) to an intuitive choice based on partial knowledge, through to a secure representation and understanding of the target word. In future studies, therefore, a schema such as that outlined by Pearson et al. (2007) or Christ (2011), whereby children are asked to define words and their responses are coded, could be used to provide a more sensitive measure of gains in understanding

We speculate that children who received the current intervention are likely to have made improvements in both the depth of their understanding of words taught and their ability to use these words in expressive language, since the programme was designed to improve both of these skills. The effects of the programme, and other similar programmes, on such measures should be examined in future studies.

Generalization

While we have demonstrated significant improvements in knowledge of the words directly taught in the programme as we anticipated this did not generalize to untaught words. The lack of generalization is as expected: our programme was short-lived and did not include teaching strategies (related to morphological understanding for example) that might support generalization.

Generalization is a major issue for vocabulary intervention studies. While a minority have found some evidence of children learning words more effectively beyond the intervention programme (e.g. Clarke et al., 2010), the majority only show intervention-related effects (Marulis and Neuman, 2010). This has implications for the number of words taught; if children only learn the target words in intervention programmes, it is arguable that such programmes should focus on teaching a larger number of words in less depth, as advocated by Biemiller and Slonim (2001). However, this involves sacrificing deep instruction and the corresponding gains in higher-order processing (McKeown & Beck, 2014). Thus a balance must be struck between teaching words in depth and teaching enough words such that children with lower vocabulary levels can reach the level of the their more able peers.

Another aspect of generalization is highlighted by the review of Christ and Wang (2011). They cite Eller, Pappas and Brown (1988), who measured children's knowledge of target words on a continuum, with generalization of a word to other contexts being the highest level of knowledge that a child can display. Therefore, a vocabulary intervention programme's effect on generalization could also be measured by children's use of taught words in other contexts, rather than their learning of untaught words. The scope of this study did not allow for measurement of this aspect of the children's learning, but future studies could consider presenting the children with target words in multiple linguistic contexts and asking them to explain the word to evaluate the depth of word learning achieved. Alternatively, child, teacher or parent reports could be used to assess the effect of the programme on the participants' use of target words in their everyday language.

The Regression Discontinuity Design

This study has confirmed that the RDD can be useful in educational settings because it allows treatment to be given to all children who need it, as determined by a specified cut-off point. Although, the RDD has lower statistical power than an RCT, our results demonstrate that, given a moderate sample size and reasonable treatment duration, the RDD is a suitable way of measuring the effects of an educational intervention. It appears that the RDD might be particularly suitable as a means of evaluating various types of educational "enrichment" programmes, where (as in the current study) an intervention can be delivered to low-performing students, preferably in group settings.

Limitations

The vocabulary programme evaluated here was manualized, but it was delivered by two different trainers meaning that there could have been small differences in delivery between the two trainers. Since the pre-tests and post-tests were administered by the same people who had delivered the interventions there is the possibility of a degree of bias in the posttest assessments. In addition our measures of the effects of the programme did not include standardized measures making it hard to characterize the vocabulary levels of the children and the absolute levels of gains achieved. A further limitation was the short duration of the programme. Future work should explore the effects of longer duration programmes on vocabulary knowledge and seek methods of instruction that yield generalized (rather than word specific) increases in vocabulary knowledge.

Formatted: Small caps

Conclusions

This study has shown that a 10-week vocabulary teaching programme can produce increases in the vocabulary knowledge of children with poor vocabulary skills. In line with work discussed in earlier, the results from our programme suggest

that combining direct instruction with interactive activities is an effective method for teaching new words (Christ & Wang, 2011; Marulis & Neuman, 2010; Steele et al., 2013; Storkel et al., 2016). However, future work needs to explore ways of producing gains that generalize to untaught vocabulary and the use of treated words beyond assessment. We believe one promising strategy may be to combine morphological training with direct vocabulary teaching, since training children to understand the morphological structure of words provides them with a strategy that can be applied to understanding the meanings of novel words (Bowers & Kirby, 2009). Methodologically, the study provides further evidence for the usefulness of the Regression Discontinuity Design as a way of evaluating educational interventions.

Acknowledgements

Funding for this project was jointly provided by Dr Jonathan Solity of Optima Psychology and the University College London Impact Studentship Scheme. We would like to acknowledge Lydia Paget's hard work in helping with teaching and collecting data as part of the project. Finally, thanks go to all the teachers and pupils who participated in this research so willingly and enthusiastically.

References

BAYDAR, N., KÜNTAY, A. C., YAGMURLU, B., AYDEMIR, N., CANKAYA, D., GÖKSEN, F. and CEMALCILAR, Z. ,2014, "It Takes a Village" to Support the Vocabulary Development of Children With Multiple Risk Factors. *Developmental Psychology,* 50(4), 1014-1025.

BECK, I. L. and MCKEOWN, M. G. ,2007, Increasing Young Low-Income Children's Oral Vocabulary Repertoires through Rich and Focused Instruction. *The Elementary School Journal,* 107(3), 251-271.

BECK, I. L., MCKEOWN, M. G. and KUCAN, L. ,2002, *Bringing Words to Life: Robust Vocabulary Instruction*. (New York: Guilford Press).

BIEMILLER, A., 2000, Teaching Vocabulary: Early, Direct and Sequential. *Perspectives: International Dyslexia Association Newsletter,* 26.

BIEMILLER, A., 2003, Vocabulary: Needed If More Children Are to Read Well. *Reading Psychology,* 24(3-4), 323-335.

BIEMILLER, A. and BOOTE, C., 2006, An Effective Method for Building Meaning Vocabulary in Primary Grades. *Journal of Educational Psychology,* 98(1), 44-62. doi: 10.1037/0022-0663.98.1.44

BIEMILLER, A. and SLONIM, N., 2001, Estimating root word vocabulary growth in normative and advantaged populations: Evidence for a common sequence of vocabulary acquisition. *Journal of Educational Psychology,* 93(3), 498-520.

BIRD, H., FRANKLIN, S. and HOWARD, D., 2001, Age of acquisition and imageability ratings for a large set of words, including verbs and function words. *Behaviour Research Methods, Instruments and Computers,* 33(1), 73-79.

BISHOP, D. V. and SNOWLING, M. J., 2004, Developmental dyslexia and specific language impairment: same or different? *Psychological Bulletin,* 130(6), 858-886.

Bishop D.VM., Snowling, M.J., Thompson, P.a. and Greenhalgh, T and the CATALISE-2 consortiu. (2017). Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. Journal of Child Psychology and Psychiatry, 58, 1068-1080. DOI: 10.1111/jcpp.12721

BLACHOWICZ, C. L. Z., FISHER, P. J. L., OGLE, D. and WATTS-TAFFE, S., 2006, Vocabulary: Questions from the classroom. *Reading Research Quarterly,* 41(4), 524-539.

BLOOM, P., 2000, *How Children Learn the Meaning of Words* (Massachusetts Institute of Technology).

BOOTE, C., 2006, Vocabulary: Reasons to Teach It, an Effective Teaching Method, and Words Worth Teaching. *New England Reading Assocation Journal,* 42(2), 24-28.

BOWERS, P. N. and KIRBY, J. R., 2009, Effects of morphological instruction on vocabulary acquisition. *Reading and Writing,* 23(5), 515-537.

CAIN, K., and OAKHILL, J., 2011, Matthew effects in young readers: reading comprehension and reading experience aid vocabulary development. *Journal of Learning Disabilities,* 44(5), 431-443.

CAPPELLERI, J., and TROCHIM, W., 2003., Cutoff designs. In S.-C. Chow (ed), *Encyclopedia of biopharmaceutical statistics* (New York: Marcel Dekker, Inc.), pp. 263–269.

CHRIST, T., 2011, Moving Past "Right" or "Wrong" Toward a Continuum of Young Children's Semantic Knowledge. *Journal of Literacy Research,* 43(2), 130-158.

CHRIST, T. and WANG, X. C., 2011, Closing the Vocabulary Gap?: A Review of Research on Early Childhood Vocabulary Practices. *Reading Psychology,* 32(5), 426-458.

Clarke, P., Snowling, M., Truelove, E. & Hulme, C. (2010). Ameliorating children's reading comprehension difficulties: A randomised controlled trial. Psychological Science, 21, 1106-1116

COYNE, M. D., MCCOACH, D. B., LOFTUS, S., ZIPOLI, R. and KAPP, S., 2009, Direct Vocabulary Instruction in Kindergarten: Teaching for Breadth verses Depth. *The Elementary School Journal,* 110(1), 1-18.

DOCKRELL, J. E., BRAISBY, N. and BEST, R. M., 2007, Children's acquisition of science terms: Simple exposure is insufficient. *Learning and Instruction,* 17(6), 577-594.

ELLER, R., PAPPAS, C. and BROWN, E., 1988, The lexical development of kindergarteners: Learning from written context. *Journal of Reading Behavior* 20(1), 5–23.

FERNALD, A., MARCHMAN, V. A. WEISLEDER, A., 2013, SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science,* 16(2), 234-248.

Goldberger, A.S. (1972). Selection bias in evaluating treatment effects: Some formal illustrations. (Discussion Paper No. 123) Madison: Unviersity of Wisconsin, Instutute of Research on Poverty.

HART, B. and RISLEY, T. R.,1995, *Meaningful Differences in the Everyday Experience of Young American Children*, 4th ed., (Baltimore: Brookes Publishing).

JUSTICE, L.M., SCHMITT, M.B., MURPHY, K.A., PRATT, A. and BIANCONE, T., 2014, The 'robustness' of vocabulary intervention in the public schools: targets and techniques employed in speech–language therapy. *International Journal of Language and Communication Disorders,* 49(3), 288-303.

KIRSCHNER, P. A., SWELLER, J. and CLARKE, R. E., 2006, Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist,* 41(2), 75-86.

LEE, J., 2011, Size matters: Early vocabulary as a predictor of language and literacy competence. *Applied Psycholinguistics,* 32(1), 69-92.

Lee, W and Munk, T. (2008) Using regression discontinuity design for programme evaluation. Journal of Statistical Methods, 21, 257-299.

LINVER, M. R., BROOKS-GUNN, J. and KOHEN, D. E., 2002, Family processes as pathways from income to young children's development. *Dev Psychol,* 38(5), 719-734.

LUDWIG, J., and MILLER, D. L., 2007, Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design. *The Quarterly Journal of Economics,* 122(1), 159-208.

MARCHMAN, V. A. and FERNALD, A., 2008, Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science,* 11(3), F9-16.

MARULIS, L. M. and 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.

MCKEOWN, M. G. and BECK, I. L., 2014, Effects of vocabulary instruction on measures of language processing: Comparing two approaches. *Early Childhood Research Quarterly,* 29(4), 520-530.

MUTER, V., HULME, C., SNOWLING, M. J. and STEVENSON, J., 2004, Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: evidence from a longitudinal study. *Developmental Psychology,* 40(5).

NAGY, W. E., 2005, Why vocabulary instruction needs to be long-term and comprehensive. In E. H. Hiebert & M. L. Kamil, eds., *Teaching and learning vocabulary: Bringing research to practice* (Mahwah, NJ: Lawrence Erlbaum.), pp. 27-44.

NAGY, W. E. and HERMAN, P. A., 1987, *Breadth and depth of vocabulary knowledge: Implications for acquisition and instruction* (Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc).

OUELLETTE, 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.

PEARSON, P. D., HIEBERT, E. H. and KAMIL, M. L., 2007, Vocabulary assessment: What we know and what we need to learn. *Reading Research Quarterly,* 42(2), 282-296. doi: 10.1598/rrq.42.2.4

RICE, M. L. and HOFFMAN, L., 2015, Predicting Vocabulary Growth in Children With and Without Specific Language Impairment: A Longitudinal Study From 2;6 to 21 Years of Age. *Journal of Speech Language and Hearing Research,* 58(2), 345.

ROMANO, E., BABCHISHIN, L., PAGANI, L. S. and KOHEN, D., 2010, School readiness and later achievement: replication and extension using a nationwide Canadian survey. *Developmental Psychology,* 46(5), 995-1007.

ROWE, M. L., 2012, A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development,* 83(5), 1762-1774.

SCOTT, J. A., MILLER, T. F. and FLINSPACH, S. L., 2012, Developing word consciousness: Lessons from highly diverse fourth-grade classrooms. In J. F. Baumann & E. J. Kame'enui (eds.), *Vocabulary instruction: From Research to Practice* (New York: Guilford Press), pp. 123-139.

SCOTT, J. A. and NAGY, W. E., 2004, Developing Word Consciousness. In E. J. Kame'enui & J. F. Baumann (eds.), *Vocabulary Instruction: Research to Practice* (New York: Guilford Press), pp. 201-217.

STANOVICH, K. E., 1986, Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly.*

TOMBLIN, J. B., RECORDS, N. L., BUCKWALTER, P., ZHANG, X., SMITH, E. and O'BRIEN, M., 1997, Prevalence of Specific Language Impairment in Kindergarten Children. *Journal of Speech, Language, and Hearing Research,* 40, 1245-1260.

Trochim, W. M. K., 2006,. Research Methods Knowledge Base [online] Available: http://www.socialresearchmethods.net/kb/ [accessed 6 December 2016].

TUCKWILLER, E. D., PULLEN, P. C. and COYNE, M. D., 2010, The Use of the Regression Discontinuity Design in Tiered Intervention Research: A Pilot Study Exploring Vocabulary Instruction for At-Risk Kindergarteners. *Learning Disabilities Research and Practice,* 25(3), 137-150.

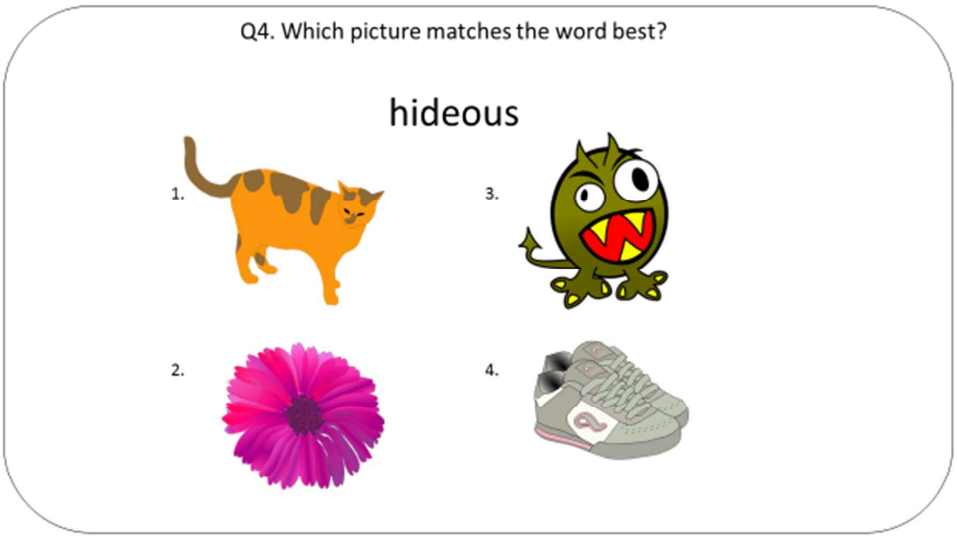
WEIZMAN, Z. O. and SNOW, C. E., 2001, Lexical Input as Related to Children's Vocabulary Acquisition: Effects of Sophisticated Exposure and Support for Meaning. *Developmental Psychology,* 37(265-279).

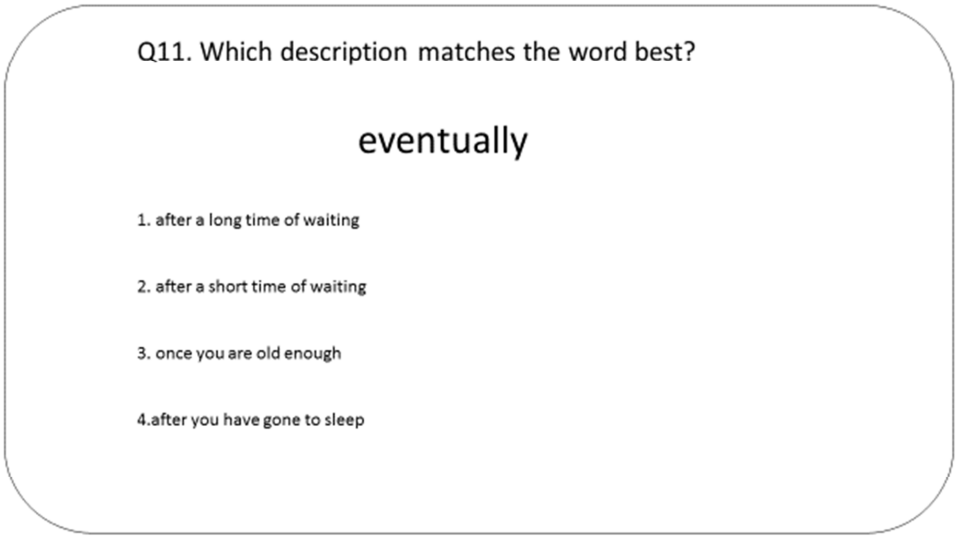
EFFECTIVENESS OF A VOCABULARY PROGRAMME

Appendices

Images used in the appendices are downloaded from www.openclipart.org and are therefore free from copyright restrictions.

Appendix 1: Sample of Word-Set Tests





EFFECTIVENESS OF A VOCABULARY PROGRAMME

Appendix 2: Example Teaching Materials

DEFINITIONS, PARAGRAPHS AND SENTENCES

DEFINITION

Fragile: Something which is fragile is easy to break. It might be made of glass or china. A person can also be described as fragile if they are feeling unwell, tired or upset.

PARAGRAPH

It was summertime, and Wayne the Weasel and his whole family had to move to a new house.

"Be careful with that box of china plates, because they're very **fragile**," said Wayne's mum as he carried them out of the door to the van. Wayne's arms were aching, and he gripped the box more tightly.

"Here, let me help you," said Dad. Wayne gave the heavy box to him with gratitude. "That's better," he said, stretching his arms as he got into the van.

ALTERNATIVE SENTENCES

Fragile: "I am feeling fragile today—I think I have the flu coming on," said mum.

The fragile glass horse shattered into tiny pieces when I dropped it on the floor.

WORD GAMES - FRAGILE

Activity 1: Which of the following objects are fragile? Call out "fragile" if you think they are.

A wine glass	A plastic bottle	A computer screen
A photo frame	A hat and scarf	A mobile phone
A brick	A teacher's chair	

Would you feel "fragile" in these situations? Call out "fragile" if you think you would.

You have a very bad cold.	You haven't slept all night.
You're going to a football match.	It's a bright, sunny day.
You got cold and wet on the way home from school.	

Activity 2: Can you think of any more objects that are fragile? Can you think of any more situations in which you might feel fragile?

EFFECTIVENESS OF A VOCABULARY PROGRAMME

WORD MAP WORKSHEET

Write today's word here:

.....

Which picture matches the word **fragile** best?



On a separate piece of paper, draw a picture of another object you think is **fragile**. What makes it **fragile**?

Which of these words means the same as **fragile**?

lovely *breakable* *easy*

Can you think of any more words that mean the same as **fragile**?

.....

Which of these words means the opposite of **fragile**?

sturdy *shocking* *sad*

Can you think of any more words that mean the opposite of **fragile**?

.....