An Evaluation of a Semi-Intensive Reading Programme

John Shillaw

It has been suggested that improvements in reading are only possible if learners read large quantities of text. This study aims to examine some of the claims that have been made.

INTRODUCTION

Despite a greater emphasis in the past 25 years on the development of oral/aural skills, the development of reading skills has retained a central role in world-wide ESL/EFL education. This is largely because the permanent, accessible nature of written texts facilitates the contextualized study of grammar and vocabulary. Certainly, as far as the latter is concerned, there is overwhelming evidence from studies of L1 English acquisition that lexical development is dependent very much on the contextualized learning of vocabulary items from reading texts (McKeowen & Curtis, 1987; Nagy, et al., 1985). L2 lexical acquisition research tends to confirm the same findings (Coady, et al., 1993; Coady & Huckin, 1997; Huckin, et al., 1993). Krashen (1989) argues strongly from the experience of L1 acquisition that the only way that L2 learners need to acquire a large number of words to read competently in English. The only way to do this, Krashen claims, is for learners to read large quantities of text and he emphasises the importance of repeated exposure to language in a variety of meaningful contexts.

Like other places in the world, English teaching in Japan has moved towards a more oral centred curriculum at the junior-high school level, but it is undeniable that the study of English in high-schools and colleges is still very much text-centred. *Chobun dokai*—reading and comprehending texts—and *yaku doku*—text translation—have dominated Japanese foreign language education since the Meiji era (Hino, 1988), primarily because the vast majority of English examinations that are taken for university entrance tend to focus on testing these two skills.

It is debatable whether the practice of *chobun dokai* and *yaku doku* and the intensive study of texts are effective in developing the top-down skills that make up reading proficiency. The approach may still, however, be efficacious in the acquisition of grammar and lexical knowledge and the development of bottom-up processes. Yet, if we contrast the arguments put forward by Krashen with the

Shillaw, J. (1999). An Evaluation of a Semi-Intensive Reading Programme.
In *The Web of English Curriculum Development* (pp.79-90).

approach to reading in Japan, it is difficult to contemplate how the task of translating what amounts to a relatively small number of texts can provide sufficient input for Japanese learners to maximise the grammatical and lexical knowledge needed to read effectively in English.

Given the situation described above, as well as the pressure of time in high-school education in Japan, it is understandable that extensive reading has remained a low-priority. The following study, I hope, I will provide some evidence of the benefit of such an experience. The study considers the impact of a controlled, semi-extensive reading scheme on two intact classes of students who attended reading classes I taught in the Foreign Language Center, University of Tsukuba during the academic year 1996-97. The next section explains the background to the study and describes the two groups who took part. Then, in the next section, I present the results of several test that were used to measure improvement of several reading-related skills. The discussion section goes on to probe the key findings and is followed by a brief conclusion.

BACKGROUND

The small scale study described in this paper investigated the improvement in grammar and vocabulary knowledge of a class of freshman Jinbun (Humanities) and Shizen (Natural Science) students after a semi-intensive reading course over the course of a full academic year. Common to both classes was the requirement for each student to read 11 simplified readers over the course of the academic year – about one book every two weeks. The readers used were predominantly from the *Oxford Bookworms* series. The books are graded from Level 1 to 6 in order of increasing difficulty as determined by text length, syntactic complexity and lexical variation. The level 1 stories are based around 700 headwords while level 6 texts contain up to 2,500 headwords.

Students were given Information about each book which included the level; the genre; the total number of words (tokens) and the number of different words (types), as well as a very brief synopsis of the story. An example from a Level 5 story is presented below. Most students said that they found the information to be very helpful when selecting a book.

LEVEL	TITLE	TOTAL WORDS	DIFFERENT WORDS	STYLE
5	THE DEAD OF JERICHO	24,498	1,702	murder-mystery
A police inspector investigates some strange murders in Oxford, England				

The results of the placement test, taken at the start of the academic year, showed that the Jinbun students were considerably better than the Shizen group in terms of reading skills. To confirm this, and to help direct students to the level reader they should begin with, the classes were administered a set of short cloze tests (Test A) developed by the University of Edinburgh (Criper, 1980) to evaluate the effect of extended reading schemes (Hill, 1992). A similar test was administered at the end of the course (Test B) to measure progress. The results of the comparison are discussed in a later section.

Students were free to select any book that they wanted, as long as it was not more than one level below the last story they had read. Students could progress to the next level whenever they felt ready. As students read a story they were instructed to write a summary of each chapter on one page of a B5 notebook. On the facing page, they were asked to record any comments they had about what they had read, to note any unknown words which they thought might be worth remembering and write a definition of it in either English or Japanese. The students were advised to be selective about the words they chose, and not to try record every one. As well as providing a convenient way of monitoring individual reading progress, it was hoped that the written tasks would stimulate students to process the texts in some depth (Craik & Tulvig, 1975), without encouraging word by word translation.

Both groups were allowed to read their selected readers in alternate classes. All the other classes for Shizen students were taken up with reading materials produced by the SRA (Parker, 1989) for young L1 readers (around Grade 3). The texts are short (200 – 400 words), graded passages and are accompanied by a variety of short comprehension questions and vocabulary exercises based directly on the texts. There are 12 texts per level and are a mixture of fiction and non-fiction stories. All the Shizen students began at the simplest level and were allowed to progress to the next after successfully completing the exercises from 4 texts with an accuracy of 80% or better. The rate of progress over the year varied considerably with the better students reading on average around 3 texts per class, while others barely read two. Reading speed appeared to be the single greatest predictor of success, but a small minority appeared to suffer from a lack of motivation to read.

The remaining classes for the Jinbun students were devoted to reading authentic texts, mostly newspaper and magazine articles. The focus of these classes was to develop an awareness of the discourse structure of the different genres and to encourage students to employ top-down processing skills. Very little attention was paid to points of grammar. Instead, students were encouraged to practice and use different vocabulary learning strategies.

RESULTS

Table 1 below shows a simple measure of progress by comparing the level of texts the students were reading at the beginning and near the end of the year.

Table 1 Readers by class and level

Jinbun	Reader Level	Frequency	Percent	Cumulative Percent
	2	2	4.7	4.7
	3	12	27.9	32.6
Beginning	4	17	39.5	72.1
	5	11	25.6	97.7
	6	1	2.3	100.0
	Reader Level	Frequency	Percent	Cumulative Percent
	3	3	6.8	6.8
	4	14	31.8	38.6
End	5	16	36.4	75.0
	6	10	22.7	97.7
	6+	1	2.3	100.0
Shizen	Reader Level	Frequency	Percent	Cumulative Percent
	1	5	15.6	15.6
Beginning	2	14	43.8	59.4
	3	13	40.6	100.0
	Reader Level	Frequency	Percent	Cumulative Percent
	2	4	12.5	12.5
End	3	14	43.8	56.3
	4	13	40.6	96.9
	5	1	3.1	100.0

The pattern is very clear with most students moving up only one level. Very few students went up two or more levels. In effect, most of the students in both classes read all of the stories at the level they started and the level above. It would appear that students in both classes took the path of least resistance and read the easiest and shortest stories which, in turn, reduced the required summary work. It should be noted, however, that when the data was collected most students should have been reading their eleventh and final story, but some were as many as 4 stories behind the target. Ultimately, all Shizen students reached level 3 or above, while no Jinbun student was lower than level 4.

As mentioned above, students took two tests made up of short cloze passages at the beginning (Test A) and end (Test B) of the reading course. The cloze procedure has been claimed to be a useful and reliable measure of global language proficiency

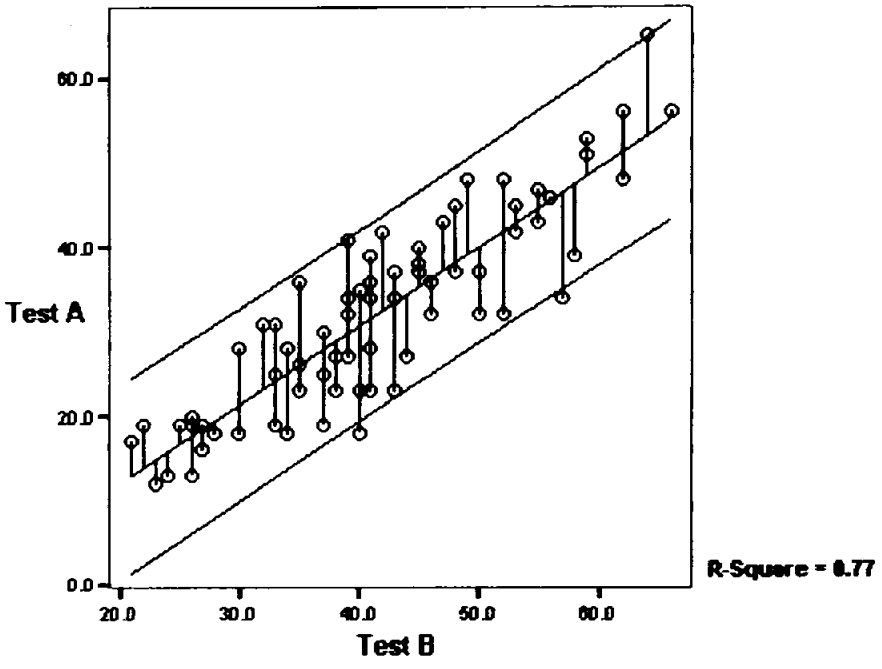
(Oller, 1979), but is generally assumed to be most effective at predicting reading, grammar and vocabulary ability. To test this assumption, the scores from both test were correlated with scores from the 6 sections of the *Eigo Kentei Shiken* (EKS) which the classes took within a week of Test B. Because the number of items on the two cloze tests were different, the test scores were standardised (Criper, 1980). The results showed that all sections of the EKS correlated significantly with both cloze tests, but that the reading section was the best predictor of performance on both cloze tests – r=0.72 and r=0.73 respectively. The scores from the error-selection part of the EKS also correlated quite strongly with the results from two cloze tests – r=0.64 in both cases. The relationship between the vocabulary test scores from the EKS and both cloze tests was around 0.5; a rather moderate correlation.

From the results, I would argue that the two cloze tests are good predictors of reading skills and grammar knowledge, but not particularly sensitive to vocabulary knowledge. This is perhaps not surprising when we consider that a large majority of the deletions on both tests were function and discourse-related words. Few lexical items were deleted from the tests, and of these, most were nouns.

The correlation between the cloze tests themselves was very strong – r=0.88, which is equivalent to an overlap in variance (R^2) of 77%. This figure can be interpreted to be a particularly strong test-retest reliability. To illustrate this point, Figure 1 on the next page displays a scatterplot of the scores from the two tests. The centre line is the regression line and the two lines parallel to it represent a 95% confidence interval for the scores. It can be observed that all the scores are quite close to the regression line and are within or about the 95% confidence interval.

Figure 1.

Relationship between Test A and Test B



Now that the reliability of the two cloze tests has been established, we can look for evidence of improvement in scores over the course of the year. It should be noted that because the scores were standardised to a common scale, they can be compared directly. Figure 2 on the next page shows that both classes did improve their scores, but the difference in mean scores for the Jinbun and Shizen classes were quite considerable. T-tests showed the difference to be significant for both Test A and B: t= 9.909, $p<0.0001$ and t= 7.664 , $p<0.0001$ respectively.

When we consider the magnitude of the gains, however, the Shizen students average improvement was 55% of the original mean compared to a gain of only 21% for the Jinbun group. Figure 3 provides more detail of the gain score distribution for the two classes, and clearly shows more Shizen students than Jinbun students at the higher end of the scale.

Figure 2.

Cloze test scores

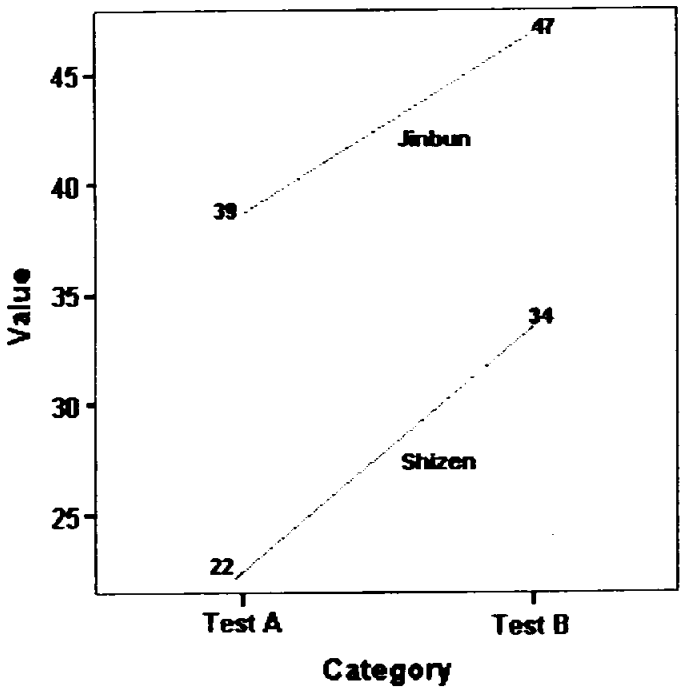
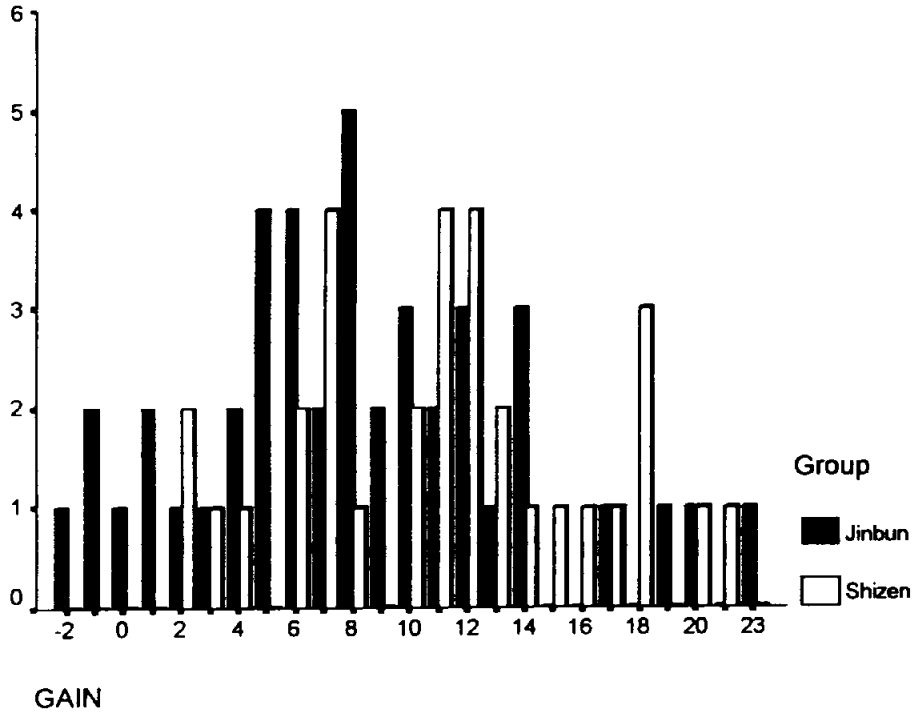


Figure 3.

Distribution of gain scores for each class



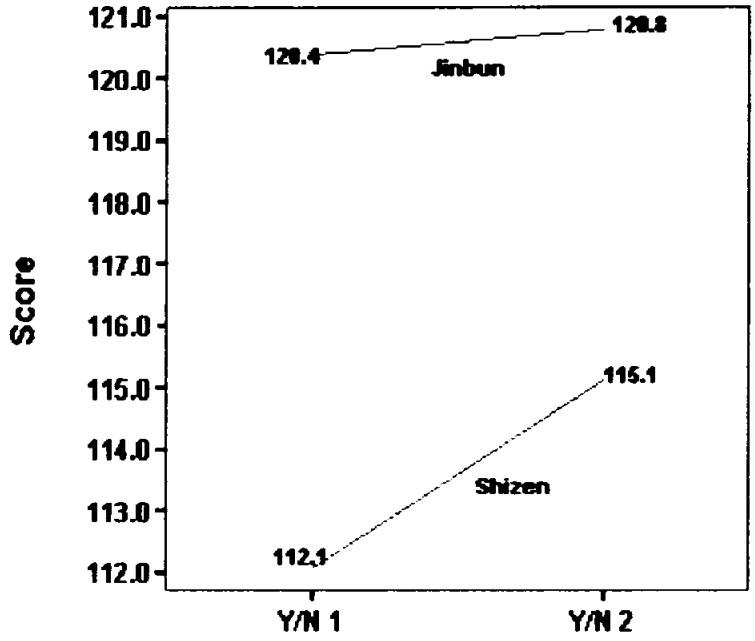
I argued earlier that the two cloze tests largely measure reading skills and grammar knowledge and were not very reliable at assessing lexical competence. Fortunately, I was able to obtain a precise estimate of vocabulary development by using data from another study I was conducting at the time (Shillaw, 1999). The research I was doing employed two Yes/No vocabulary tests to measure gains in vocabulary size over the last 5 months of the course. Conventionally, Yes/No tests are a word list which combine a set of real words that are representative of the words in English which learners need to learn (e.g., the first 10,000 words), with a set of nonwords—words that resemble legitimate words of English but do not exist. Learners check which words they know and their vocabulary size is estimated by applying a formula which adjusts the proportion of real words known according to the proportion of nonwords identified as known. Yes/No tests have been successfully used with young L1 subjects (Anderson & Freebody, 1983; White, et al., 1989) and L2 learners (Meara & Buxton, 1987; Meara & Jones, 1988; Meara, et al., 1994).

I created the two Yes/No tests mentioned above from a random, stratified sample of the 3,272 most frequent nouns occurring in the 100 million word British National corpus (Leech, 1993). Unlike the research cited above, I used only real words in the tests, having found in earlier studies (Shillaw, 1999) that using nonwords to correct vocabulary size estimates was neither effective nor reliable. The approach I took was based on the application of the Rasch model (Rasch, 1960; Hambleton, et al., 1991), which offered a more reliable and tractable model for vocabulary size estimation. A full description of the method is beyond the scope of this paper and can be found in Shillaw (1999).

The mean scores from the two Yes/No tests are presented in Figure 4 on the following page. These are not the exact scores from the tests, but are standardised measures of ability where 100 represents the population average. What we observe is that the Jinbun students show superior vocabulary knowledge on both Yes/No tests, but that their lexical development is minimal over the period of 5 months. In contrast, the Shizen students make quite good gains and close the gap on the Jinbun group. In more tangible terms, the scores translate into an average gain of 172 words for Shizen students and 40 words for Jinbun students. This is equivalent to Shizen students learning around 9 words per week compared to 2 words for the average Jinbun student. The gains for the Shizen students are very similar to those reported in a study of lower-intermediate level Japanese high-school and university students using a different testing method (Schmitt & Meara, 1997).

Figure 4.

Yes/No test scores



DISCUSSION

Overall, the figures from the analyses presented in the previous section present a positive picture of improvement in reading, grammar and vocabulary skills for both groups of students. The gains made the Shizen students are particularly impressive. But to what extent can we attribute the gains to the impact of the semi-intensive reading experience? Clearly, in the absence of a control group, it is not possible to draw any conclusions about causality, but we can perhaps make a few tentative observations from a comparison of the performance of the two groups who were involved in the study.

The Shizen students clearly benefited most from the reading scheme. There are two possible reasons why this may be so. Firstly, their knowledge of grammar and vocabulary, and their competence in reading improved because the materials were more suited to learners at their level of proficiency. Possibly, the materials contained little to benefit the Jinbun students. Secondly, the fact that the Shizen students read a greater variety of texts than the Jinbun group might have maintained their interest and motivation to read.

My first suggestion is supported by Krashen's claims (Krashen, 1989) that learners will only benefit from the reading experience when the materials are fairly

closely matched to their proficiency level. This was certainly the case for the Shizen group and for most of the Jinbun students at the beginning of the course. However, as the course progressed, the simplified readers may not have provided the same level of stimulation for an increasing number of Jinbun students. This had been considered and was one of the reasons why more authentic materials were used in the second part of the course. There is no way to determine how much the students benefited from reading authentic texts, but it may have been less than was anticipated. Possibly, this may have been because the Jinbun students were not quite ready to handle the kind of lexically based tasks I had set. This hypothesis is given some support by the estimated average vocabulary size of the Jinbun group, which was around 2,5000 headwords by the end of the course. It has been estimated (Laufer, 1992; Coady, et al., 1993) that learners of English need around 3,000 headwords for an adequate comprehension of authentic reading materials and it is possible this shortfall of 500 words might be critical in blocking access to such texts. If this is the case, the implication is that the Jinbun students and learners who are in a similar situation to them, are trapped somewhere in between the upper limits of vocabulary coverage in simplified readers and the lexical requirement of non-simplified texts.

Turning now to my second point, maintaining the motivation to read so many texts was a problem for both groups, but particularly the Jinbun students. From an early stage in the course, it was clear that a majority of the class were reluctant to invest very much time in reading. Trying to get the students to produce summaries of what they had read was especially difficult. The reason for the poor motivation may have been related to the point I made above about the level of the texts. In addition, the topics of the stories at the higher levels do not appear to have been as interesting as the readers at the lower levels. However, my feeling is that there were additional reasons related to a poor class dynamic which was evident from the start of the year, a point confirmed by the teacher who taught them writing.

The Jinbun students never interacted in the supportive way that the Shizen students did, who, as a group, were quite happy to discuss the stories they had read and to compare notes on the new vocabulary they had found from their reading. While they remained more positive than the Jinbun class about the reading experience, they did complain—albeit good-naturedly—that reading and summarising the simplified readers was very time-consuming. The strongest resistance to reading was observed in the SRA classes, which some students saw as an opportunity to do as little as possible. Not surprisingly, this particular group of

students performed poorly on the cloze tests and showed little evidence of gains in grammatical or lexical knowledge.

CONCLUSION

Overall, the outcome of the study reported here provides some evidence in support of the effectiveness of semi-extensive reading as a way of improving reading skills. The fact that the work was on a small-scale obviously limits the generalizability of the findings, and future studies will need to refine the rather simple methods and measures adopted here. In my opinion, the most important consideration for further research is to establish a causal link between the experience of extensive reading and gains in the component skills that make up reading proficiency. This study concentrated on the skills essential to bottom-up processing—grammar and vocabulary knowledge—at the expense of higher level, top-down skills. This discrepancy clearly needs to be addressed.

Future studies will also need to be concerned with questions of motivation and interest and how they relate to achievement. The evidence presented above is largely anecdotal, so there is a clear need to utilise one or several qualitative approaches to explore the dynamic between reading improvement and the affective domain.

References

Anderson, R. C., & Freebody, P. (1983). Reading comprehension and the assessment and acquisition of word knowledge. In B. Hutson (Ed.), *Advances in reading/language research* (pp. 231-256). Greenwich, CT: JAI Press.

Coady, J., & Huckin, T., (Eds.). (1997). *Second language vocabulary acquisition.* Cambridge: Cambridge University Press.

Coady, J., Magoto, J., Hubbard, P., Graney, J., & Mokhtari, K. (1993). High frequency vocabulary and reading proficiency in ESL readers. In T. Huckin, M. Haynes & J. Coady (Eds.), *Second language reading and vocabulary learning* (1993). Norwood, N. J.: Ablex.

Craik, F. I. M., & Tulvig, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of experimental psychology, 104*, 268-284.

Criper, C. (1980). *EPER Placement/Progress Test A.* [IALS]. Edinburgh: University of Edinburgh.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory.* Newbury Park: Sage.

Hill, D. R. (1992). *The EPER (Edinburgh Project on Extensive Reading) guide to organising programmes of extensive reading.* [IALS]. Edinburgh: University of Edinburgh.

Hino, N. (1988). Yakudoku: Japan's dominant tradition in foreign language learning. *JALT Journal, 10* (1&2), 45-55.

Huckin, T., Haynes, M., & Coady, J. (1993). *Second language reading and vocabulary*. New Jersey: Ablex.

Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *The Modern Language Journal, 73* (4), 440-464.

Laufer, B. (1992). How much lexis is necessary for reading comprehension? In P. Arnaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 126-132). London: Macmillan.

Leech, G. (1993). 100 million words of English. *English Today, 9* (1), 9-16.

McKeowen, M. G., & Curtis, M. E., (Eds.). (1987). *The nature of vocabulary acquisition*. Hillsdale: Lawrence Erlbaum.

Meara, P., & Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language testing, 4*, 142-154.

Meara, P., & Jones, G. (1988). Vocabulary size as a placement indicator. In P. Grunwell (Ed.), *Applied linguistics in society* (pp. 80-87). London: CILT.

Meara, P., Lightbown, P. M., & Halter, R., H. (1994). The effect of cognates on the applicability of YES/NO vocabulary tests. *The Canadian modern language review, 50*(2), 296-311.

Nagy, W., Herman, P., & Anderson, R. C. (1985). The acquisition of word meaning from context by children of high and low ability. *Reading research quarterly, 20*, 233-253.

Oller, J. W. (1979). *Language tests at school: A pragmatic approach*. London: Longman.

Oxford Bookworms Series. Oxford: Oxford University Press.

Parker, D. H. (1989). *Teacher's handbook, SRA reading laboratory 2b*. Science Research Associates.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danmarks Paedagosike Institut.

Schmitt, N., & Meara, P. (1997). Researching vocabulary through a word knowledge framework. *Studies in second language acquisition, 19*, 17-35.

Shillaw, J. (1999). *The application of the Rasch model to Yes/No vocabulary tests*. Unpublished PhD dissertation. Wales, Swansea.

White, T. G., Slater, W. H., & Graves, M. F. (1989). Yes/No method of vocabulary assessment: valid for whom and useful for what? In S. McCormick & J. Zutell (Eds.), *Cognitive and social perspectives for literacy research and instruction* (pp. 391-397). 38th Yearbook National Reading Conference, Chicago.