

# A Trial Study of Integrating a Systematic Extensive Reading Component With a Proscribed Non-ER Syllabus

Ron Murphy • Ashley Peace

## INTRODUCTION

When extensive reading (ER) is discussed in pedagogical terms the scenario typically involves a dedicated reading-centered class where class time is devoted to working with activities involving graded reader texts. These situations are ones in which the ER component is central to the lesson plan or the course entirely. But in cases where the syllabus or curriculum is strictly set without an ER component - such as curriculums with specified learning aims, common textbooks, and common tests - using graded readers systematically in-class or for homework becomes a challenge. In such a situation, finding occasional opportunities to insert ER-related activities is easy. But the benefits of ER are achieved through long-term *extensive* reading, so the challenge in such a situation is to create a methodology in which ER can supplement and complement the requirements of the syllabus while still allowing students the reading exposure necessary for long-term gains.

This paper reports on a classroom study which trialed an ER program consisting of graded reader-related activities, word-count reading goals, and quantitative assessment mechanisms in an oral communication class. There were four aims for this exploratory research. The first was to determine if ER can be incorporated into an oral communication class on a regular basis without compromising the learning aims of the course. The

second aim was to determine if a 50,000 word-per-semester reading goal is reasonable and what effects such a goal would have on students' reading comprehension. The third aim was to test a cloze procedure as a way of measuring reading comprehension gains. The final aim was to determine if a non-ER control group achieved incidental reading comprehension gains after one semester of normal (non-ER) coursework.

## THE RESEARCH

The two keys of extensive reading involve reading extensively and reading material that is easily comprehensible. (Day & Bamford, 1998). Nation (2009) and others recommend learners read 300,000–500,000 words per year and to continue this over several years for gains in reading and vocabulary acquisition to be durable. In cases where these reading goals were accomplished, improvements in various reading measures as well as acquired vocabulary have been well documented (in Mason & Krashen, 1997).

Various assessment methods were used to measure the gains mentioned above. However, most of the gains mentioned above occurred in ER-friendly environments unconstrained by a proscribed syllabus lacking an ER component. An ER program that complements a proscribed syllabus throughout a semester can be, in and of itself, pedagogically valid but certain measures of students' achievements would lend further validity to such a program.

In assessing students' reading level or comprehension, cloze procedures are particularly effective in ER environments. Cloze procedures have existed since Taylor first introduced them in 1953, originally as a method for measuring readability of texts. While critics claim that cloze tests measure primarily local and grammatical comprehension, a wealth of studies (Jonz, 1990; Brown 1980; Anderson, 1972; Alderson, 1979; Oller and Conrad, 1973) have concluded that cloze procedures are valid and accurate measures of general reading comprehension.

## THE STUDY

This study took place in the first semester of the 2013–2014 academic year at a national university in Western Japan among mostly first-year non-English majors in an oral communication course with curriculum-wide specified learning aims, a common textbook, and a common final exam. First-year classes accounted for two experimental groups (Aer=20, Ber=19) and two control groups (Ac=18, Bc1=25). A third control group (Bc2=15) was comprised of second-year upper-level students enrolled in an advanced course. When planning the study, the authors (identified as A and B) attempted -- according to their experience and knowledge of which faculty groups tended to be stronger or weaker in English -- to assign the presumed higher and lower-ability groups as control and experimental groups. This was entirely a guess, as there were no data in which to make an analytical judgment. In the end, the groups' reading levels were found to be almost identical, as the mid-semester TOEIC Bridge scores (most students took that test in week 8 of the semester) showed (Table 1). The only group that was predictably distinguishable from the others was B's second-year advanced group (Bc2).

**Table 1. Reading Comprehension Gains on Cloze Tests**

	<u>Temple</u>		TOEIC		<u>Ski</u>		Faculty
	<u>Pre</u>	<u>Post</u>	<u>Bridge*</u>	<u>Pre</u>	<u>Post</u>	<u>Bridge*</u>	
A's ER	1.0	1.3	78	1.4	1.6	78	Humanities
A's control	0.8	1.2	76	1.1	1.4	75	Sci/Engineer
B's ER	0.9	1.0	77	1.1	1.2	70	Elec Engin
B's control	0.9	1.0	77	1.3	1.4	77	Humanities
B's Adv. control	1.2	1.4	83	1.4	1.5	85	Pro Course

\* Reading score only.

Two cloze tests were administered. The readability of the two texts were confirmed via the Range and Frequency Test (Nation, 2005), which found that 90 percent of the

words in the two texts fell within the first 2,000 words of the General Services Word List (West, 1953). Precursors of these two tests had been used in a similar study (the ‘Sam’ study - results yet unpublished) at the same school the previous semester. After the results of the Sam study were analyzed, the tests were revised and shortened for the Ashley study. Pre-tests were administered to all groups in the first week of the semester. Each group was divided, with one-half of the students receiving the *Temple* cloze text and the other half receiving the *Ski* cloze text. The same groups were later administered the same texts as a post-test in the 13<sup>th</sup> or 14<sup>th</sup> week of the semester. Reading comprehension gains or losses were derived from the results of these tests.

Also in the first week, the two experimental groups were introduced to graded reader usage and the ER component of the class, which included the rationale behind extensive reading and an emphasis on reading regularly and at a level that is easy for the students. This component required students to read at least 50,000 words from graded readers in the English Center’s Graded Reader Library (GRL). Given the constraints on the student and teacher by the set-syllabus as defined above, the authors considered 50,000 words an appropriate minimum goal. Five one-thousand- word graded readers are within reach of most of our students. Ten graded readers of that length in one 15-week semester was deemed reasonable. Extra points were awarded to students for each subsequent 5000-word increment that they passed. For each book read, students were required to pass a short quiz associated with the book on the MReader website ([www.mreader.org](http://www.mreader.org)). This website provides quizzes for many graded readers and it also provides a way for teachers to track students’ reading progress. Students who pass the quiz receive automatic credit for the number of words the graded reader contains. In this study, students who failed the quiz were required to write a summary report on the book to demonstrate that they had read and understood the book. Word count credit was then given.

As for the treatments given to the two types of groups, the control groups received no further treatments – they were taught in the teachers’ usual manner. The experimental groups were given in-class communicative activities associated with the graded

A Trial Study of Integrating a Systematic Extensive Reading Component With a Proscribed Non-ER Syllabus readers that comprised part of their weekly homework assignments. These activities, however, were set in the context of the learning aims specified in the syllabus. These aims included various speech acts and negotiation techniques such as topic expansion, follow-up questions, negotiating meaning, turn-taking, agreeing and disagreeing, and listener expressions.

## FINDINGS

### *Reading Comprehension Gains/Losses*

The findings reveal little overall reading comprehension gains by the end of the semester, as measured by the cloze tests. Both of A's groups gained, with the control group gaining the most - 0.4 and 0.3 percentage points on the *Temple* and *Ski* tests respectively. A's ER group gained 0.3 and 0.2 percentage points on the two tests. Each of B's groups gained 0.1 percentage point on both tests, with only the advanced *Temple* group gaining more (0.2). The findings are somewhat surprising, especially among B's groups, in that greater gains would likely be expected among the ER groups rather than the control groups. In comparing A's ER results to B's, the somewhat better performance by A's ER group could be a factor of the different levels the two groups read at. As Table 2 shows, 67 percent of A's group read (their most recent book) at the easiest level, while 89 percent of B's ER group read their most recent book at one or two levels above that. However, because the TOEIC Bridge reading scores of each group are very similar, each ER group should have been reading graded readers at similar levels. Without further evidence this is more conjecture than conclusion, but it does raise two important points: that ER students should read at their comfort level and that an accurate level check should be a part of any ER program.

The gains of A's control group were the most interesting finding of this study. This group's post-test scores equaled all the other groups, but its starting level was noticeably lower. The reasons for this low starting score need to be better understood for a

**Table 2. Level of Most Recent Book Read**

<u>Levels read</u>	<u>Overall</u>	<u>A</u>	<u>B</u>
Easiest	38%	67%	5%
Red	25	5	47
Yellow	35	29	42
Green	3	0	5

conclusion to be reached. It could be that the group's gains were achieved incidentally. If so, the nature of students' starting level and the reading exposure they experienced in that semester need to be understood.

#### *ER Component Complementing Learning Aims?*

The effectiveness of integrating an ER component into an oral communication class in a manner that supported the class' required learning aims was another aim of this trial study. Could the teacher somehow exploit the graded reader homework in a way that supported these aims?

The two authors had no quantitative measure of the effects of their respective approaches, and neither author coordinated lesson plans with the other. However, they independently employed similar approaches in applying the ER elements into their respective classes. Both teachers primarily used the students' own reading homework as the basis for group discussion in English during the class. Fortunately, the learning aims for the class are not topic-specific but instead focus on basic speech acts and communicative strategies. This allows for the characters, plots, and themes of students' graded readers to provide the gist for the students' discussions within the framework of the learning aims.

Anecdotally, but significantly, both authors observed the students respond to the ER in-class activities in ways they had previously seen in the course of their teaching careers

– that students enjoyed and were stimulated by the variety of topics encountered when graded readers served as the basis for discussion in English. In the authors’ experience with this approach, they have noted that when students move from partner to partner or group to group to discuss their book and listen to others, there is an element of anticipation and energy to the discussion. Additionally, both authors have observed that students are generally more involved – as though they are more ‘vested’ in the topic – when using a graded reader story that they themselves have chosen versus a topic dictated by the textbook or teacher. While the inverse of this is sometimes true (common topics, i.e. hometown, hobbies, and holidays, require less preparation to discuss in class than does a graded reader and are thus preferred by some students in some cases), graded readers have been observed to work effectively at this university as discussion material for periods ranging from 10 to 40 minutes per class.

As Table 3 below shows, the 50,000-word goal for graded readers was achieved by nearly all students in B’s class, with about half exceeding the minimum by a few thousand words and others exceeding it by the equivalent of a few graded readers. The poor reading response of A’s students is curious. Half of the students did not meet the goal. Teacher A speculated that one reason for this was the fact that these students read mostly at the lowest level, which has books with the lowest word counts. Students might have felt that having read a book they had met their goal. What is interesting, and counter-intuitive, is that the largest gains on the cloze test are found among the students who read the *least*. No conclusion can be found by sorting the students by their TOEIC Bridge reading scores, as these are distributed randomly.

**Table 3. Total Words Read and Gain or Loss**

	<u>Teacher A</u>			<u>Teacher B</u>		
	<u>Words Read</u>	<u>Gain/Loss</u>	<u>TOEIC Bridge*</u>	<u>Words Read</u>	<u>Gain/Loss</u>	<u>TOEIC Bridge*</u>
1.	123,531	0.2	72	92,513	-0.2	80
2.	89,019	0.3	88	88,637	0.1	86
3.	67,272	0.5	82	88,320	0.0	74
4.	65,918	0.0	80	87,401	0.0	66
5.	62,815	0.2	84	81,802	-0.1	72
6.	59,970	-0.1	68	68,582	-0.2	68
7.	55,942	0.2	82	67,187	-0.2	60
8.	55,248	0.2	82	66,821	0.0	78
9.	53,712	-0.1	72	58,886	0.1	76
10.	49,288	0.1	84	58,769	-0.3	66
11.	46,407	0.3	86	57,317	0.1	72
12.	39,871	0.6	86	56,982	0.4	50
13.	38,741	0.5	78	56,515	0.3	66
14.	36,680	0.4	70	55,115	0.4	86
15.	36,665	0.2	80	53,792	0.0	82
16.	35,772	0.1	86	52,883	0.6	78
17.	30,497	-0.1	80	51,615	0.1	72
18.	26,131	0.2	60	51,357	0.4	84
19.	19,017	0.9	68	44,555	0.5	66

\* *TOEIC Bridge reading scores only.*

### *Performance of Cloze Procedure and Tests*

Another aim of this study was to determine whether a cloze procedure would work both procedurally and as an assessment mechanism for reading comprehension in the ER component of the oral communication class being trialed. Procedurally, the tests proved easy to administer. Twenty-five minutes were required to complete the testing procedure, from the initial explanation to the students prior to the pre-test to the collection of the tests after completion. Students had no trouble understanding how to take



the tests, and most had finished by the 15-minute mark. Scoring cloze tests, however, is somewhat painstaking. If using an *acceptable alternative* scoring system (Kobayashi, 2002; Abraham & Chapelle, 1992; Brown, 1980; Alderson, 1979), as the authors did in this study, a consistent approach to the many ‘judgment calls’ is critical to maintain inter-rater reliability. Scoring the Ashley study was relatively straightforward, as the authors were able to consult the answer table developed after the Sam study for a majority of the items. However, scoring decisions on the Ashley test’s new items had to be discussed among the authors, and reviving the mental approach the Sam study authors applied to the semantic subtleties of scoring that initial test was a challenge.

Although the two tests measured comparably on the Range and Frequency Test (Nation, 2005), a post-analysis of the texts’ features revealed that the *Ski* test contained many more adjective items than the *Temple* test (Table 4). This imbalance likely made the *Ski* test comparably easier than the *Temple* test because with the *acceptable alternative* scoring system, adjectives are more likely than items from other lexical groups to be accepted, thus increasing the likelihood that the *Ski* overall point total would be inflated compared to *Temple* scores.

**Table 4. Item Analysis – Types per Cloze Passage**

<u>Item Type</u>	<u>Temple</u>	<u>Ski</u>
Verb	16	14
Noun	12	10
Adjective	3	10
Adverb	3	3
Preposition	2	0
Conjunction	0	1
Total	36	38

## CONCLUSION

The findings of this study show that an extensive reading component can be incorporated into an oral communication class in a way that supports the learning aims of a curriculum-wide set-syllabus course with no ER component. Whether or not a similar ER component as described in this study could effectively fit into a different course is a matter for another study. It is safe to assume that such an ER component could fit naturally into a reading course. It is also possible to envision how using graded readers might augment a writing course.

These findings also suggest that a 50,000-word goal may or may not be reasonable. Nearly all of B's students were able to exceed it, many by as much as 50 percent and a few by approximately 90 percent. However, half of A's students failed to achieve the goal, though it is suspected that they mistook the number of books that they read, instead of the number of words that they read, as being sufficient to satisfy their ER homework assignment. Better tracking by both the teacher and students via the MReader program can avoid such problems. Regardless of what word-count goal is most appropriate, students must be sure to read extensively at an appropriate level. Whether the cloze procedure detailed in this study could serve as an assessment method for that purpose should be investigated.

These findings reveal that the cloze procedure and the particular pre and post-tests in this study could be used as part of a general ER component. However, these tests and the meaning of their results need further study. Firstly, the issue of lexical comparability between the cloze texts and tests should be investigated. Secondly, the question of whether the difficulty level of the tests needs a common reference (i.e. TOEIC) should also be studied. In conclusion, this study shows that a cloze procedure can work procedurally in an ER program, that the MReader software is an effective tracking and management system, and that an ER program such as defined in this study can effectively complement the learning aims of a communicative class even when those aims are unrelated to

A Trial Study of Integrating a Systematic Extensive Reading Component With a Proscribed Non-ER Syllabus extensive reading.

## REFERENCES

- Abraham, R.G. and C. Chapelle. 1992. The meaning of cloze test scores: An item difficulty perspective. *The Modern Language Journal*, 76(4), 468-479.
- Alderson, Charles J. 1979. The cloze procedure and proficiency in English as a foreign language. *TESOL Quarterly*, 13(2), 219-227.
- Anderson, J. 1972. Application of cloze procedure to English learned as a foreign language in the Territory of Papua and New Guinea. *English Language Teaching*.
- Brown, J. D. 1980. Relative merits of four methods for scoring cloze tests. *The Modern Language Journal*, 64(3), 311-317.
- Day, R. & J. Bamford. 1998. *Extensive reading in the second language classroom*. Cambridge University Press: Cambridge.
- Jonz, John. 1990. Another turn in the conversation: what does cloze measure? *TESOL Quarterly*, 24(1), 61-83.
- Kobayashi, M. 2002. Cloze tests revisited: Exploring item characteristics with special attention to scoring methods. *The Modern Language Journal*, 86(4), 571-586.
- Mason, B. & S. Krashen. 1997. Extensive reading in English as a foreign language. *System*, 25, 91-102.
- Nation, I.S.P. 2005. *Range and frequency program*, [http://www.vuw.ac.nz/lals/staff/Paul\\_Nation](http://www.vuw.ac.nz/lals/staff/Paul_Nation).
- Nation, I.S.P. 2009. *Teaching ESL/EFL Reading and Writing*, Routledge. New York.
- Oller, J. & C. Conrad. 1973. Cloze tests of second language proficiency and what they measure. *Language Learning*, 23, 105-118.
- Taylor, W.L. 1953. Cloze procedure: a new tool for measuring readability. *Journalism Quarterly*, 30, 415-433.
- West, M. 1953. *A general service list of English Words*. London: Longman, Green & Co.