An Investigation of the Selective Deletion Cloze Test as a Valid Measure of Grammar-Based Proficiency in Second Language Learning

Gregory S. Hadley and John E. Naaykens
An Investigation of the Selective Deletion Cloze Test as a Valid Measure of Grammar-Based Proficiency in Second Language Learning

Gregory S. Hadley and John E. Naaykens

Introduction

Few issues in the field of second language research have been as contentious as cloze testing. Over the years, opinions in the TEFL academic community have been divided over the applicability of cloze tests for the second language classroom. Some contend that cloze tests measure a language learner’s overall communicative ability in the target language (Hanania and Shikhani 1986). Others maintain that cloze tests assess only the most basic of second language learning and reading comprehension (Shanahan, Kamil and Tobin 1982). Still others support a moderate position. Ikeguchi (1995), who quotes Bachman (1990:86-89), states that cloze testing:

... hold[s] potential for measuring aspects of students' written grammatical competence, "knowledge of vocabulary, morphology, syntax, and phonology," as well as textual competence, "knowledge of cohesive and rhetorical properties of text" in second language (p. 167).

Some years earlier, Bachman (1982:61-70) reported that certain types of cloze tests, such as the selective deletion cloze, can be used to investigate a subject’s knowledge of written discourse items such as context cohesion, syntax and strategic textual comprehension. Anderson (1979) adds that cloze testing correlates more closely with grammar tests than with reading tests, and according to Bowen et al. (1985:376), the selective deletion cloze is ideal for testing vocabulary and grammar. Claims such as these should prompt us to find out for ourselves if cloze tests, such as the selective deletion cloze, can measure a subject's knowledge of grammar. Would students with higher scores on a selective deletion cloze test also score higher on a criterion-referenced examination designed to measure grammatical competency?
We will consider this question as we review a 1996 study conducted at Niigata University. The purpose of this study was to investigate whether the selective deletion cloze correlates highly with traditional, grammar-based tests. Many language teachers in the national university system opt for criterion-referenced tests (C-RTs) which attempt to measure grammatical knowledge (Garland 1996). Putting aside the issue of whether language teachers should focus primarily on grammatical proficiency, a selective-deletion cloze test, if proven to be a valid measure of grammatical competency, might provide a time-saving method of examination which is both fair to students and easier to grade for teachers. Before looking at the findings of this study, however, a brief history has been provided for those new to cloze testing.

**Close Testing: An Overview**

Cloze testing was first introduced by W.L. Taylor (1953), who developed it as a reading test for native speakers. He defined the term "cloze" from a gestalt concept which teaches that an individual will be able to complete a task only after its pattern has been discerned:

A cloze unit may be defined as: any single occurrence of a successful attempt to reproduce accurately a part deleted from a 'message' (any language product), by deciding from the context that remains, what the missing part should be (p. 416).

Cloze tests consist of a text (usually two or three paragraphs) which has had words or parts of words deleted from it. Test subjects must draw from their knowledge of the language in order to write appropriate words in the blanks (see Table One).

<table>
<thead>
<tr>
<th>Ours was the marsh (1)</th>
<th>down by the river, (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>as the river wound, (3)</td>
<td>miles of the sea. (4)</td>
</tr>
<tr>
<td>first most vivid and (5)</td>
<td>impression of the identity (6)</td>
</tr>
<tr>
<td>things, seems to me (7)</td>
<td>have been gained on (8)</td>
</tr>
<tr>
<td>memorable raw afternoon towards (9)</td>
<td></td>
</tr>
</tbody>
</table>

At such a time (10) found out for certain, (11) this bleak place overgrown (12) nettles was the churchyard; (13) that Philip Pirrip, late (14) this parish, and the (15) wife of the above, (16) dead and buried.

These are the words to choose from:

I were that My to within with a of broad twenty and Georgiana of evening country

*Table 1: Example of a Fixed-Rate Cloze Test.*
There are at least five main types of cloze tests available to language teachers: The fixed-rate deletion, the selective deletion (also known as the rational cloze), the multiple-choice cloze, the cloze elide and the C-test (Ikeguchi 1995; Weir 1990; Klein-Braley and Raatz 1984).

In the fixed-rate deletion, after one or two sentences, every nth word is deleted. Usually every fifth or seventh word is deleted, but Brown (1983) suggests that longer texts with every eleventh or fifteenth word deleted can be used with subjects who have a lower level of language proficiency. Multiple choice cloze tests provide the subjects with several possible items to choose from for each blank. The cloze elide inserts words which do not belong in the text, and requires the subjects to identify the incorrect words plus write appropriate items in their place. The C-test consists of deleting only part of every second word in a text, and asks subjects to complete each truncated word. In the selective deletion or rational cloze, the tester chooses which items he or she wishes to delete from the text. The goal for teachers using this test is not only to fine tune the level of difficulty of the text, but also to measure the knowledge of specific grammatical points and vocabulary items. Let us now consider whether the selective deletion cloze truly is a reliable measure of grammatical knowledge.

**Subjects**

One group (see Table Two) from Niigata University was selected for this study. As Table Two shows, all were native Japanese speakers consisting mostly of first year Science majors. No special criteria was used in selecting or excluding the subjects. Neither was the group tested on their English proficiency level before entering the course. However, classroom experience with the subjects led us to believe that most group members had limited speaking, listening and writing skills, typically representative of a Japanese university first year EFL class (cf. Wadden 1993).

<table>
<thead>
<tr>
<th>English 1B, Niigata University, 1996-1997</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Language</strong></td>
<td>Japanese</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td>18 (82%) 19 (18%)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td>Male (55%) Female (45%)</td>
</tr>
<tr>
<td><strong>Department</strong></td>
<td>Science (91%) Education (9%)</td>
</tr>
<tr>
<td><strong>Skill Level</strong></td>
<td>False Beginners</td>
</tr>
<tr>
<td><strong>Total Number Subjects</strong></td>
<td>22</td>
</tr>
</tbody>
</table>

Table 2

---

113
Materials

*Interchange Two* (Richards, et al. 1993) was used as the primary text. The selective deletion cloze was created from one of the general interest reading texts in the first chapter of the course book (Richards *et al.* 1993:7, see Table Three). While the subjects had read the text several months earlier, we were fairly certain that very few, if any of the students had read the text again since that time. The cloze test consisted of a 133 word passage with 25 blanks, meaning that roughly 19% of the total text was deleted. Test–retest was conducted two separate times on this particular cloze. At a probability rating at less than one percent that the results are due to chance (*p* < .01), the reliability coefficient for this cloze test reached a moderate level of significance (*r*<sub>xx</sub> = +.56 and +.60).

There are many things people remember about the sixties. Some people remember it for mini-skirts, the Beatles, hippies and the flower children. It was a time when young people "owned" the world and thought that anything was possible. In art, fashion, and music, the big names were often in their early twenties, and some of them were already millionaires! The sixties was a time when young people used to do whatever things they wanted. "Don't trust anyone over 30!" they said. In the arts, people like Andy Warhol created "pop art." And fashions changed, too. The mini-skirt became popular, and then the "unisex" look followed. Young people started wearing blue jeans everywhere -- to school, fancy restaurants, and concerts. Many of them had very long hair and wore lots of rings, beads and bracelets.

Table 3: Text Selected for Use in this Study. Adapted from Richards, *et al.* (1993:7).

Procedure

The cloze test (see Figure One) was administered to the subjects two times, separated by a period of two weeks. During the second administration, a grammar-based test created by the textbook designers was also given to the subjects (Richards, *et al.* 1993:168–172). The instructions were given to the students verbally and in written form, both in English and Japanese, to facilitate a clear understanding of the task. On each occasion, the cloze tests were collected after 20 minutes. One significant variable that was different, however, is that the first test was administered during a regular class session, while the other was given during their midterm test. While this is certainly not standard practice when studying the validity of a certain test design, allowing this procedure provided a venue to find out how the cloze test would function under a variety of classroom conditions.
An Investigation of the Selective Deletion Cloze Test as a Valid Measure of Grammar-Based Proficiency in Second Language Learning

Reading and Vocabulary

Student Name: ____________________________
Student Number: ___________________________

Instructions: Fill out the blanks below with the correct words.

There are many things people remember about the sixties. Some people _______ it for mini-skirts, _______ Beatles, hippies, _______ the flower children. It _______ a time _______ young people “owned” the _______ and thought that anything was _______. In art, fashion, and music, the big names _______ often _______ their early twenties, and some _______ them _______ already millionaires! The _______ was _______ time when young people _______ to do whatever _______ they wanted. “Don’t _______ anyone over 30!” they said. In _______ arts, people like Andy Warhol created “pop _______.” And fashions changed, too. _______ mini-skirt _______ popular, _______ then _______ “unisex” look followed. Young people started _______ blue _______ everywhere — to school, fancy _______ , and concerts. Many of them had very long hair and wore lots of rings, beads and bracelets.

Figure 1: Selective Deletion Cloze Test Designed for this Study.

Analysis

The tests were graded by two scorers. The classroom teacher graded the grammar-based tests using the key provided in the teacher’s manual (Richards, et al., 1993:189–190), while a native English speaking TEFL lecturer graded the tests using the Semantically Acceptable Word (SEMAC) Method. Typically, cloze tests can be graded using either the Exact Word or SEMAC scoring method. In the exact word method, the cloze test blanks must be completed with the exact word as was in the original text. Correct answers receive 1 point, while any other response receives no points. SEMAC scoring allows subjects to write answers which are grammatically and lexically appropriate, although not the original words deleted from the text. For the purposes of this experiment, it did not matter whether the exact word method or SEMAC method was used, since they both correlate highly with each other (cf. Owen et al. 1996; Hadley and Naaykens, in press). However, SEMAC scoring may require a subjective
judgement by the scorer. In order to avoid the cloze test scores to be influenced by personal knowledge of the subjects, an evaluator unacquainted with the subjects was chosen. Before grading the tests, the blind evaluator was given a manuscript of the complete text, and instructed to allow any words in the cloze that were either synonymous, lexically or grammatically correct. Mistakes in historical accuracy, and minor spelling errors were ignored. If it was difficult to ascertain whether an answer was acceptable or not, it was scored as incorrect.

After the scores were totaled, all of the data was analyzed using the VAR Grade for Windows 2.0 software package (Revie 1997). The method of analysis was set up as a directional one-tailed test which used the Pearson r correlation coefficient. The cloze test scores were correlated with the scores of the grammar-based test, and resulted in a correlation coefficient of + .72 (See Figure Two).

According to Brown (1993:132-141), at p .005, the critical level of significance for a group of 22 is approximately + .51 (see also Fisher and Yates, 1963). This suggests that the correlation between the grammar-based test and the selective-deletion cloze may be quite significant.

**Implications for Language Teachers**

It would be foolhardy if language teachers completely changed their testing practices simply on the basis of this one study. However, the findings of this research tends to suggest that selective-deletion cloze tests could be used in place of or alongside of grammar-based language
An Investigation of the Selective Deletion Cloze Test as a Valid Measure of Grammar-Based Proficiency in Second Language Learning
tests. If careful consideration is given to the design of the selective-deletion cloze, it has a high potential for reliability, even under less than desirable testing conditions. It may be even more reliable than tests which our learners are frequently exposed to: tests which have been thrown together late at night by language teachers under the pressure of several deadlines. Conservative use of the selective deletion cloze could provide teachers with a time-saving method of testing their learners. Learners could be assured that, despite the brevity of the test, their level of grammatical competence in the target language is being, to a certain degree, reliably measured. Both teacher and learners might then be liberated from the unnecessary amount of time normally spent on testing, and more time could be dedicated to studying the target language.

Conclusion

It is hoped that language teachers will begin experimenting with cloze testing as a viable option to the traditional tests which are normally administered in university language classrooms. Even if some are uncertain about the reliability and validity of the selective deletion cloze for use as a C-RT, it could still be used as a quick measure to see if the learners are making progress in the course.

This study opens avenues for future research. For example, to what extent would a selective-deletion cloze correlate with a test measuring oral proficiency, or with a listening proficiency test? If such scores did consistently correlate highly, would this suggest that cloze tests can measure more than just grammatical competence in second language learning? These are just a few of the many questions which deserve further investigation as we continue our search for innovative and effective methods of second language testing.

References


