

The cloze test: Do SEMAC and exact word scoring methods correlate?

Sinha Geoffrey

要 約

クローズテストは元来、人間が不完全な図形を見た時、欠けている部分を補って知覚する過程に関するゲシュタルト心理学が元になっている。クローズテストは第2言語学習者の知識を測る上で信頼できる価値のある方法と言える。この論文では1997年にOWEN氏によって行われた、正確な回答(EW word)と内容的に容認出来る回答(SEMAC)を比較したときにどちらもクローズテストの結果と強く関連づいていることを実証した。この研究では同じ文章から、意識的に異なる単語を抽出して行われたもので、横浜市の私立中学校で13歳から14歳の海外生活経験者(帰国子女)とレベルは高いけれど帰国子女ではない生徒、合計43名を2つのクラスに分けて実施された。

Table of contents	Page
Title page	15
概要 (Abstract)	15
Table of contents	15
List of tables	16
List of appendices	16
1. Introduction	17
2. The cloze test	17
3. Methodology	18
4. Results and discussion	19
5. Implications for the classroom and future research	22
6. Conclusion	22

List of tables	Page
1. Variables in cloze testing	17
2. Bachman's (1985) item classification framework	18
3. Syntactic, strategic and cohesive knowledge	19
4. Correlation between EW and SEMAC scoring for Forms A and B	19
5. Causal factors within the strategic and cohesive items	20
6. Results for strategic Item 3	20
7. Results for strategic Item 5	20
8. Results for cohesive Item 7	20
9. Results for cohesive Item 19	20
10. Common features within both tests	21
11. Results for Item 13	21

List of appendices	Page
A. The Cloze Test (Form A)	23
B. EW and SEMAC answers (Form A)	23
C. The Cloze Test (Form B)	24
D. EW and SEMAC answers (Form B)	24

1. Introduction

Cloze tests were originally based upon the principle of “closure”, which is a theory within Gestalt psychology referring to a natural predisposition towards understanding and completeness, particularly when confronted with confusion and fragmentation (Litz, D. and A.K. Smith, 2006: 2). This paper examines the claim that Exact Word (EW) and SEMantically Acceptable (SEMAC) scoring methods correlate highly (Owen, et.al, 1997: 42). Encompassed within this analysis is the investigation of a number of underlying causal relationships that were observed to assist in explaining the surface fact of correlation. To obtain the results of this experiment, a single passage was rationally deleted to create two essentially distinct cloze tests. These tests were then administered to two mixed level groups of thirteen to fourteen year old students who have either lived overseas for an extended period of time before returning to live in Japan (returnee learners) or who are high level non-returnee learners (total = 43), at a private junior high school in Yokohama, Japan, where the author is employed. The next section introduces some fundamental cloze testing variables, and provides a contextual base from which this experiment is approached. Section Three describes the methodological processes, and provides a referential taxonomy of syntactic, strategic and cohesive knowledge. In Section Four the experiment’s data is analyzed, and relationships between the EW and SEMAC scoring methods and correlation are observed and discussed. Section Five outlines various implications for the classroom environment, and for the continuation of future research. This assignment is concluded in Section Six.

2. The cloze test

This section first of all illustrates a number of cloze testing variables that are discussed within this paper, and then provides a contextual background to the cloze test, before concluding with a discussion of what cloze tests measure.

2.1 Background

Since the earliest employment of cloze tests in the field of Teaching English as a Foreign Language (TEFL), a large number of variables affecting cloze tests have been documented. Table 1 below, illustrates the variables that are addressed in this paper.

Table 1: Variables in cloze testing

Variable	Description	Source
Scoring method	EW: Only exact word answers are accepted SEMAC: Exact word and semantically acceptable answers are accepted.	Alderson (1979: 219-20); Oller (1972: 151-2); Kobayashi (2002: 575)
Deletion method (a)	Fixed rate ratio: Words are deleted at a fixed ratio of every <i>n</i> th word. Rational deletion: Words are deleted according to a pre-determined rationale (e.g. syntactic, strategic or cohesive items).	Alderson (1979: 219); Bachman (1982: 66; 1985); Hughes, 2003:189)
Deletion method (b)	The 2 nd half of every word is deleted (C-test).	Klein-Braley and Raatz, (1984: 136);
Item categorization	A classification framework for identifying the amount of context needed to answer a blank.	Bachman (1982/5); Jonz (1990); Kobayashi (2002)
Choice of text	A rationale for the passage employed.	Hughes (2003: 193);
Test takers’ proficiency	Discusses the relationship between second language (L2) proficiency and test takers’ results.	Alderson (1979); Jonz (1987)
Test management	Includes the administration of the test and test rubrics.	Brown (2004); Hughes (2003: 215)

Note: Other variables such as the starting point of deletion (Alderson, 1979); the frequency that a deleted item occurs in the text (Brown, 1989) and calculating text difficulty by statistical formula (Stubbs and Tucker, 1974) are not included in this assignment.

Cloze tests were first employed in the 1950s to measure native English speakers’ ‘readability of passages’ (Oller and Conrad, 1971: 183). However, their efficacy as a ‘powerful and economical measure of English-language proficiency for non-native speakers’ (Stubbs and Tucker, 1974: 241) was recognized by at least the early 1970s (Oller, 1972: 151; Alderson, 1979: 219; Bachman, 1982/5). The construction of the cloze test follows a simple procedure of deleting a number of words, in this particular case 25, from an appropriate passage for testing. Research findings posit this ease of construction on the one hand, with a high return of reliability and validity on the other (Jonz, 1990: 72; Stubbs and Tucker, 1974: 239). Indeed, in many countries and institutions today, the cloze test is ‘viewed as a guarantee of quality’ (Klein-Braley and Raatz, 1984: 135).

2.2 What do cloze tests measure?

Although researchers’ opinion on this matter is divided, Litz and Smith (2006) suggest that at a broad level cloze tests test ‘the total communicative effect of discourse and the underlying linguistic competence’ (2006: 1). More specifically, cloze items necessitate different types of knowledge and different levels of context and proficiency to complete. Bachman’s (1985) framework, illustrated in Table 2 below, provides an excellent starting point for understanding these different types and levels.

Table 2: Bachman's (1985) item classification framework

Type	Title	Description
1	<i>Within clause</i>	Test takers employ clausal knowledge to fill the blank.
2	<i>Across clause, within sentence</i>	Test takers employ inter-clausal knowledge to the sentence boundary.
3	<i>Across sentence, within text</i>	Test takers' broaden their scope to the sentential and inter-sentential levels.
4	<i>Extra-textual</i>	Test takers utilize knowledge from outside the text.

Clearly, items at the first level or at lower inter-sentential levels are the easiest type of deletion to answer, because they do not measure reading beyond the clausal level. This suggests that test takers need only be concerned with 'the immediate environment of a blank' (1979: 225). However, research has also shown that cloze tests require comprehension at the second and third 'sentential or suprasentential levels' (Saito, 2003: 39), and a number of researchers concur that some cloze tests necessitate an understanding of fourth level higher integrative and global language skills (Bachman, 1982/5; Brown, 2001: 393; Jonz, 1990: 70; Oller, 1972: 157).

3. Methodology

This section describes methodological processes and components within this study; pilot studies, setting, participants, materials, test construction and administration.

3.1 Pilot studies

According to Oppenheim (1992) 'every aspect of a survey has to be tried out beforehand to make sure that it works out as intended' (1992: 47). Dornyei (1994) recommends that researchers perform at least an initial and final piloting study to ensure the early recognition and resolution of 'problems and potential pitfalls' (1994: 64-5). Due to its small-scale nature and time restrictions, this particular survey consisted of two initial stage pilot studies. In line with Converse and Presser (1986), these pilot studies were conducted with the assistance of 'that familiar source of forced labor – colleagues ... [and friends]' (1986: 53, as cited in Dornyei, 1994: 66). These volunteers provided invaluable help in compiling a list of acceptable answers for the SEMAC answer sheet.

3.2 Materials

The passage utilized in this study is 310 words in length, and was taken from Wikipedia, an Internet-based encyclopedia that is continually updated by everyday internet users. A fundamental and deliberate intention of this paper was to utilize the same

passage to create two distinctive tests; one of syntactic/strategic knowledge (Form A), the other of syntactic/cohesive knowledge (Form B). It was conjectured that utilizing the same passage would reduce extraneous variables, and more clearly highlight any differences observed between these knowledge types. Its choice was further inspired by its cultural and topical relevance to teenagers living in Japan. Sasaki (2000) found that '[c]ultural schema' (2000: 85) (i.e. the inclusion of content that test takers can relate to) had a positive effect on the number of answers that that test takers attempted. A recent study in a Japanese national newspaper showed that cell-phone ownership for Japanese aged from 13 to 16 years was at a national high of 98% (The Japan Times, 2008: 3). In addition, television commercials depicting well-known and popular Japanese celebrities using "cool" cell phones with built-in MP3 players and digital music files suggests that participants already possess extensive cell phone knowledge. Thus, a high level of 'real-world knowledge' (Bachman and Palmer, 1996) is contained within this passage (The passage and EW answers are supplied in Appendix A (Form A) and Appendix C (Form B)).

3.3 Setting and participants

3.3.1 Setting

The test was conducted during the last week of lessons before winter vacation. This setting is significant because test results could no longer affect learners' overall end-of-semester results. It is likely that this had a negative affect on test-takers' extrinsic motivation to complete the test (other than for research purposes).

3.3.2 Participants

Participants were Japanese males, aged from 13 to 14 years ($n = 43$). Almost half of the participants are returnee learners, and have attained native or near-native levels of English proficiency. The remaining participants have studied English formally for at least two years, and have a high level of non-native speaker proficiency for the amount of time that they have studied, though not as high as the returnee students. Participants were first ranked from highest to lowest according to the results from their last end of semester (in-school criterion tested) exam results and were then divided according to odd and even numbers. Significantly, participants' levels were sub-divided into two proficiency groupings (returnee and non-returnee), which provided an unexpected opportunity to test the efficacy of cloze tests within different proficiency levels.

3.5 Administration and scoring

3.5.1 Administration

The tests were administered at the beginning of class. Students were informed that usual testing protocol applied (i.e. a time limit of twenty minutes was allocated for completing the test, while talking, dictionaries and other study aids were prohibited). Before the test, learners were instructed to go through the passage quickly and to answer the items that they knew first, before taking time to go through the passage more carefully for the more difficult items.

3.5.2 Scoring

Both tests were first scored in accordance with the EW key, and then with the SEMAC key. One point was awarded for each correct answer. Misspelled items were awarded points if they were understandable and correct according to the scoring key. Completed papers were sorted from highest to lowest according to EW results (The EW and SEMAC answer key is supplied in Appendix B (Form A), and Appendix D (Form B)).

3.5.3 Test construction and knowledge types

In order to test whether different types of knowledge affect scoring correlation, two tests were constructed by rational deletion from one passage. Rational deletion enables the test maker to ‘control, by design, the abilities which a given cloze test measures’ (Bachman, 1985: 535-6). The first test (Form A) contained five syntactic and twenty strategic deletions. The second test (Form B) contained five syntactic and twenty cohesive deletions (see Appendix A). Table 3 provides a taxonomy of syntactic, strategic and cohesive knowledge types.

Table 3: Syntactic, strategic and cohesive knowledge

Type	Description	Source(s)
Syntactic	- Is comprised largely of grammatical items, which are contained within the clause	Richards and Schmidt, 2002: 535
	- Consists of the lowest (and easiest to answer) level	
	- Example: ...while riding a bus OR train...	
Strategic	- Is ‘compensatory’ in nature.	Skehan, 2003: 158
	- Corresponds with Bachman’s Types 1 and 4 categories	Bachman, 1982: ##
	- Example: ...buy anything from MASCARA to jet planes...	Jonz, 1990: 77
Cohesive	- Consists of words that ‘mark’ grammatical and lexical relationships across clauses and/or sentences	Bachman, L. and A.S. Palmer, 2004: 68
	- Encompasses short and long-range lexical items.	Skehan, 2003: 158
	- Corresponds with Type 2 and 3 categories	Jonz, 1990: 77
	- Includes collocation, reiteration, conjunction, subordinators and referential determiners.	
	- Example: JAPAN was the first country to...	

4. Results and Discussion

In this section the statistical analysis and quantitative results gathered from Forms A and B are presented in tabular format and are discussed in relation to scoring correlation. Calculations were first made by hand using the formula, $\rho = 1 - 6\sum (a-b)^2 / n(n^2 - 1)$ (Owen et.al, 1997: 99), and then were computer checked using SPSS (version 14.0). Spearman’s ρ was used because it is more reliable than Pearson’s Product Movement (PPM) for less than thirty observations (ibid: 98/9).

Table 4: Correlation between EW and SEMAC scoring for Forms A and B

Form	Participants (n=)	Difference	Difference Squared	Correlation Spearman’s (ρ)	Reliability Cronbach’s (α)
A	22	10	58.5	0.953	0.895
B	21	0	35.5	0.977	0.945

Significance was set at p<0.5.

Table 3 illustrates that the correlation for EW and SEMAC scoring were very high. However, Kenny (as cited in Bachman, 1990: 260) observes that correlations are ‘not ... inferential’, meaning that the underlying causes of correlation must be explained, and then linked to the wider research body, to further research into cloze-testing.

4.1 EW and SEMAC scoring correlation

Research into cloze testing has found that different deletion rates will create distinctive tests of knowledge and proficiency (Alderson, 1979: 221). Unfortunately, the only conclusions that can be drawn in this case are that correlation between Form A and Form B proved very consistent, and consistently very high. Clearly, correlation occurs when test takers’ language ability manifests itself as consistent test results. This finding poses a number of questions: (1) How did this consistency replicate itself not once, but twice in two tests? (2) Is it simply because EW and SEMAC scoring will always correlate highly? (3) If so, what factors cause such highly recurrent correlations? (4) Can they be positively identified, and do they exist within these tests?

4.1.1 Distinct features of the strategic and cohesive tests

In this sub-section, correlation and its positive or negative causal connections with regard to four observations from Forms A and B are examined in the light of the above questions. Table 7 illustrates these features and highlights whether the feature contributed positively (✓), neutrally (△) or negatively (×) to correlation.

Table 5: Causal factors within the strategic and cohesive items

Form	Description	√ / △ / ×
A	- Contained a high number of EW-and-SEMAC answers.	×
	- Participants did not appear to use cultural schema.	△
B	- Contained a high number of EW-only answers.	√
	- Participants appeared to use background knowledge and cultural schema.	×

The first feature related to the characteristics of the deleted items, and to the proportions of available EW-only and EW-and-SEMAC answers. Form A consisted of 16% more EW-and-SEMAC answers than Form B. Strategic items simply generated more EW-and-SEMAC combinations because of their characteristic ‘compensatory’ nature (Canale and Swain, 1980, as cited in Bachman, 2004: 99). Brown (2001) observes that good language learners use ‘strategies to fill in gaps in their own competence’ (2001: 209). For example, a good language learner with a limited vocabulary will endeavor to find as many meaningful contexts for the words that he or she knows. This is further exemplified in Table 6 below. Mascara (Item 3) is a small and inexpensive object (especially when compared with a jet plane), but in reality test takers’ are displaying an acceptable level of semantic understanding by writing the name of virtually any small, inexpensive shopping product.

Table 6: Results for Strategic Item 3

Item	EW correct	SEMAC correct
3 ...credit cards to buy anything from MASCARA to jet planes...	0 / 0%	21 / 95%

When the SEMAC scoring method was employed for Item 3, 21 correct answers (a 95% increase) were recorded. Certainly, this phenomenon reduces the amount of correlation among EW and SEMAC scoring methods by widening the gap between test takers scores for both data sets. Nevertheless, while Item 3 highlights the logic of the SEMAC scoring method, i-mode (Item 5) in Table 7 below is more interesting in that it does not appear to be overtly difficult, but caused great difficulty among test takers.

Table 7: Results for Strategic Item 5

Item	EW correct	SEMAC correct
3 ...keitai can be connected to the internet through services such as I-MODE.	0 / 0%	3 / 12%

Sasaki (2000) found that test takers and test makers often categorize items very differently. Thus, although Bachman (1985: 540) categorizes Item 5 as a Type 4 item, it is likely that learners did not investigate enough of the context beyond the blank for clues, and hence classified Item 5 as Category 1. In addition, the fact that test takers possessed a high amount of ‘[c]ultural schema’ (Sasaki, 2000: 85), and yet still recorded many incorrect or missing answers for Item 5 is perplexing. Regardless, missing answers have neither a positive or negative affect upon the calculations for correlation. In comparison, cohesive items are characterized by the way that they meaningfully ‘mark’ (Bachman and Palmer, 2004: 68) inter-clausal and inter-sentential relationships, and cohesive items are characteristically EW-only. Japan (Item 7) in Table 8 below, exemplifies this; it is EW-only, and it is marked (reiterated) seven times across the passage, including in the title.

Table 8: Results for Cohesive Item 7

Item	EW correct	SEMAC correct
A JAPAN was the first country where users...	10 / 48%	n/a

Although the reliance across large textual boundaries may cause difficulties for these test takers, the high number of EW-only items means that test takers can perform well, on the condition that they are alert to this feature. The paradox of cohesive items positively contributes to high correlation, because EW-only items will not widen the gap between scoring methods; and likewise, items that many test takers answer wrongly will, again, merely add strings of zeros to the data.

Table 9: Results for Cohesive Item 19

Item	EW correct	SEMAC correct
19 ≥ the low per-message <i>cost</i> , the ability to	0 / 0%	3 / 14%

Enhance (Item 19) in Table 9 above is an example of the tendency to create (or in this case disrupt) cross-clausal and sentential relationships by pragmatic combinations with other words. Item 19 is medially located in the sentence, and it is also embedded between two more semantically dependent items. This deleted sequence caused many problems for the test takers; no more than 14% were able to receive a point for this item.

4.1.2 Common features within both tests

Could correlation also be the result of certain common features that are extant within both tests? In this

sub-section, four common traits are examined to see if they lead to correlation. Table 4 illustrates some fundamental considerations.

Table 10: Common features within both tests

Proficiency level
- Within each group existed two sub-levels (native/near-native and high-level learners).
Test takers' performance
- Test takers performed better with EW-only items than with EW/SEMAC items.
Missing answers
- A high number of missing answers were present within Forms A and B, and most missing answers occurred within the latter part of each test.
Item characteristics
- Forms A and B both contained items that tested syntactic knowledge.

The first observation regards two distinct levels of proficiency within both groups, presenting an opportunity to examine whether EW and SEMAC scoring methods in particular favor one level of proficiency over another. On average, lower level test takers, who scored under 12 points by EW scoring method, increased by only 14% with SEMAC scoring. In comparison, higher test takers, who scored above 12 points by the EW method, increased their score with the SEMAC method by 26%. Therefore, learners who scored highly according to the EW method also scored highly with the SEMAC method. On the surface this means that these scoring methods do favor higher level learners. However, this result could also be interpreted as a positive correlation between test takers (whether high or low level) who attempted more answers and “high” achievement. It seems obvious in retrospect that participants who attempted more answers had more chances of receiving correct answers than participants who left answers blank, but this data went unnoticed for some time. It also seems likely that factors such as test takers’ attitudes to the tests and their awareness of good testing strategies are as important, if not then perhaps more important than language proficiency alone. Secondly, test takers performed better when answering EW-only items (total=53% success rate) when compared with the alternative EW-and-SEMAC items (total=41% success rate). Perhaps test takers’ preference for searching within the clause and utilizing only lower level knowledge (Categories 1 and 2) to complete the test is the reason for their success rate with EW-only blanks. This indicates (if it is true) that test takers are not aware of higher levels that can be utilized during testing situations, indicating that they are more likely to leave Category 3 and 4 items blank or answer them incorrectly. Certainly, the reason for the high proportion of EW-only items is not related to item

characteristics: EW-only items where characterized by nouns, proper nouns, verbs and prepositions. The EW-and-SEMAC category contained only a slightly wider range: Nouns, verbs, determiners, conjunctions, an article and surprisingly, a proper noun.

Thirdly, a large number of missing answers were discovered throughout the data, particularly towards the end of both tests. Sasaki utilized ‘missing’ (2000: 95) as a category within her research of the cloze test, and concluded that missing answers were the result of unfamiliar content. The high number of missing answers is therefore surprising. Why were 63% of missing answers grouped towards the latter half of these tests? Three possibilities (or a combination of them) are likely. First of all, perhaps test takers were not allocated enough time to complete the test, or secondly, the latter section was more difficult than the earlier section, leading to the third possibility that they effectively gave-up after an initial effort.

The concluding observation concerns the inclusion of five very specific syntactic test items within Forms A and B. Syntactic items, usually the easiest type of cloze test item, were initially included as a control item to compare the results of the strategic and cohesive test items. However, an unforeseen situation was noted relating to an innate difficulty that is particular to Japanese, and almost certainly encompasses all levels of learner in Japan. This situation was noted by Kobayashi (2002):

Articles present particular difficulty, possibly because Japanese learners do not have articles in their mother tongue and, therefore, do not realize that articles are necessary when they are missing’ (2002: 577)

Three article-related errors are presented in Table 5, below.

Table 11: Results for Item 13

		No. of respondents (#)	% of total
Form B (13)	¥≥ to talk on A keitai...’ (correct)	5	20%
	¥≥ to talk on THE keitai...’ (incorrect)	5	28%
	¥≥ to talk on ___ keitai...’ (missing)	13	52%

48% of test takers correctly identified the need for an article to fill the blank, but unfortunately selected the wrong one. Indeed, sometimes “a” and “the” can be interchangeable, but not in this case, because the passage refers to “keitai” in general. 52% of learners left this item blank. This would have been a correct

answer, if not for the test stipulation that all answers were to be filled.

This sub-section has addressed the question of correlation by illustrating how common traits within Forms A and B are related to the high correlations in this paper. It was found that the presence of articles in the tests, and the presence of high category items contributed positively towards correlation, because these factors made the tests more difficult for test takers. Further, missing answers do not radically alter correlation. Finally, cloze tests do not conclusively favor one proficiency level over another, although test takers' attitudes may be a significant factor within correlation.

5. Implications for the classroom and future research

The employment of this cloze test to native/near native and high level learners allowed the identification of more general and also more specific areas of concern with regard to test takers' 'readability of passages' (Oller and Conrad, 1971: 183). Participants' reading tasks could be refined to extend their foci beyond the immediate word, clause and sentence, and to be aware of mid and long range textual relationships within reading passages. This could also extend to a Consciousness-Raising (C-R) approach to addressing specific grammatical difficulties (e.g. articles).

Correlation indicates that both scoring systems are 'measuring the same quality' (Owen, 1997: 42), however, it would be interesting to undertake an extended comparative case-study to observe how changes within underlying qualities (such as reducing the amount of missing answers) affect this correlation. In addition, enlarging the populations within these tests would allow the application of different statistical formula. Certainly, the incorporation of interviews to gather empirical qualitative evidence would support and enhance the statistical data, and remove the researcher's need to conjecture about the causes for test takers' results. Further, one disadvantage of Bachman's (1985) framework is that it is not always 'clear which category [each cloze item] should go in' (Brazil, 1995: 15). Therefore, the construction of a modified framework that is neither too general, nor too restrictive, yet is suitable for different genres would be a challenging future research goal.

6. Conclusion

In conclusion, the claim that EW and SEMAC scoring methods correspond highly was found to be accurate.

Analyses of two cloze tests, which were constructed to test syntactic, strategic and cohesive knowledge types revealed a range of variables. These were observed to contribute positively, neutrally or negatively towards correlation. It was found that differences between Forms A and B contributed less positively towards correlation than the similarities. It was also observed that test takers performed better with EW-only items, which largely belonged to lower level categories, but less well with EW-and-SEMAC items, which were largely higher level. Finally, this experiment has raised the author's awareness not only of the statistical processes inherent in testing, but also of the complex variables that underlie one of the most simple classroom tests.

References

- Alderson, J.C. (1979) 'The Cloze Procedure and Proficiency in English as a Foreign Language' TESOL Quarterly, Vol. 13 # 2: 219-227
- Bachman, L.F. (1982) 'The Trait Structure of Cloze Test Scores' TESOL Quarterly, Vol. 16, # 1: 61-70
- Bachman, L.F. (1985) 'Performance on Cloze Tests with Fixed-Ratio and Rational Deletions' TESOL Quarterly, Vol. 19, # 3: 535-556
- Bachman, L.F. (1990) *Fundamental Considerations in Language Testing* Oxford: Oxford University Press
- Bachman, L.F. and A.S. Palmer (2004) *Language Testing in Practice* Oxford: Oxford University Press
- Brazil, D. (1995) *Classroom and Spoken Discourse* Birmingham, UK: University of Birmingham
- Brown, D.H. (2004) *Language Assessment: Principles and Classroom Practices* New York: Pearson Education
- Brown, D.H. (2001) *Teaching by Principles: An Interactive Approach to Language Pedagogy* New York: Pearson Education
- Dornyei, Z. (2003) *Questionnaires and Second Language Research: Construction, Administration and Processing* London: LEA Publishers
- Jonz, J. (1990) 'Another Turn in the Conversation: What

Does Cloze Measure?' TESOL Quarterly, Vol. 24, # 1: 61-83

Klein-Braley, C. and U. Raatz (1984) 'A survey of research on the C-Test' Language Testing, Vol. 1, # 2: 134-146

Kobayashi, M (2002) 'Cloze Tests Revisited: Exploring Item Characteristics with Special Attention to Scoring Methods' The Modern language Journal, Vol. 86: 571-586

Litz, D.R. and A.K. Smith (2006) 'Semantically Acceptable Scoring Procedures (SEMAC) Versus Exact Replacement Scoring Methods (ERS) For "Cloze" Tests: A Case Study' UAE University

Oller, J.W. (1972) 'Scoring Methods and Difficulty Levels for Cloze Test of Proficiency in English as a Second Language' The Modern Language Journal, Vol. 56 # 3: 151-158

Oppenheim, A.N. (1992) Questionnaire Design, Interviewing and Attitude Measurement London: Pinter

Owen, C., Rees, J., Wisener, S. and P. Crompton (1997) Testing Birmingham, UK: University of Birmingham

Sasaki, M. (2000) 'Effects of cultural schemata on students' test-taking processes for cloze tests: a multiple data source project' Language Testing, Vol. 17: 85-111

Skehan, P. (2003) A Cognitive Approach to Language Learning Oxford: Oxford University Press

SPSS Incorporated (2005) SPSS Advanced Statistics 14.0 Chicago, IL: SPSS

Stubbs, J.B. and G.R. Tucker (1974) 'The Cloze Test as a Measure of English Language Proficiency' The Modern language Journal, Vol. 58 # 5/6: 239-241

Appendix A. The Cloze Test (Form A)

Technological advances of the cell phone in Japan

All basic-model keitai include an extensive phone/address book, alarm clock and stopwatch and games, plus varying degrees of image enhancement. Most keitai that were sold within the last five years are capable of taking high-quality still and video images. These images can be embedded in messages and sent to other keitai.

Newer models allow the user to listen to music, or (1.) _____ movies and television. There (2.) _____ even versions that can be used as debit or credit cards to buy anything from (3.) _____ to jet planes. Most keitai can be connected (4.) _____ the internet through services such as (5.) _____. Japan was the first country where users (6.) _____ browse text-only (7.) _____ sites, and many Japanese sites have sub-(8.) _____ designed especially (9.) _____ keitai users. (10.) _____ of the most popular services allows users to check (11.) _____ schedules, the local (12.) _____ forecast, or their daily (13.) _____.

(14.) _____ is frowned upon to talk (15.) _____ a keitai while riding a bus (16.) _____ train, so most users use email, (17.) _____ messages by email, or 'text-(18.) _____ is a simple, fast and private form of (19.) _____. This combined with (20.) _____ low per-message cost, the (21.) _____ to enhance messages with (22.) _____ characters, emoticons, pictures and small animations, and to write in English or Japanese, has made (23.) _____-messaging extremely popular in Japan among users from every generation. (24.) _____ are particularly fond of this (25.) _____ of communication, and many schools ban the use of keitai on campus.

The inclusion of so many capabilities into a single hand-held device is without doubt a remarkable feat of modern technology. The necessity of these features is another issue, but for many people in Japan a keitai is something that they cannot live without.

- | | |
|-----|---------------|
| 1. | watch |
| 2. | are |
| 3. | mascara |
| 4. | to |
| 5. | i-mode |
| 6. | can |
| 7. | internet |
| 8. | sites |
| 9. | for |
| 10. | one |
| 11. | train |
| 12. | weather |
| 13. | horoscope |
| 14. | It |
| 15. | on |
| 16. | or |
| 17. | sending |
| 18. | messaging |
| 19. | communication |
| 20. | the |
| 21. | ability |
| 22. | special |
| 23. | text |
| 24. | teenagers |
| 25. | form |

Appendix B: EW and SEMAC answers (Form A)

EW	SEMAC
1. watch	see, enjoy, download, share
2. are	
3. mascara	
4. to	with
5. i-mode	NTT DoCoMo
6. can	could, might
7. internet	net, web, www, computer
8. sites	homepages, urls, websites
9. for	
10. One	
11. train	their, one's, T.V., movie, bus
12. weather	
13. horoscope	stars, blog, games, website
14. It	
15. on	using, with
16. or	
17. Sending	Texting, Exchanging, Thumbing
18. messaging	communication, -ing
19. communication	communicating, talking, speaking, email
20. the	a
21. ability	capacity, option
22. special	personal, funny, anime, cartoon
23. text	e-mail, mail, keitai
24. Teenagers	Students, Girls, Boys, Children
25. form	method, kind, type, way

Appendix C: The Cloze Test (Form B)

Technological advances of the cell phone in Japan

All basic-model keitai include an extensive phone/address book, alarm clock and stopwatch, games plus varying degrees of image enhancement, and most keitai sold within the last five years are capable of taking high-quality still and video images. These images can be embedded in messages and sent to other keitai.

Newer (1.) _____ allow the user to listen to music, and watch movies and television. (2.) _____ are even versions that (3.) _____ be used as debit or credit (4.) _____ to buy anything from mascara to jet planes. (5.) _____ keitai can be connected to the internet (6.) _____ services such as i-mode. (7.) _____ was the first country where users can (8.) _____ text-only internet sites, (9.) _____ many Japanese sites have sub-sites designed especially for keitai users. One of the most popular (10.) _____ allows users to check train schedules, (11.) _____ local weather forecast, or (12.) _____ daily horoscope.

It is frowned upon to talk on (13.) _____ keitai while riding a bus or train, (14.) _____ most users use email. Sending messages by (15.) _____ or 'text-messaging', (16.) _____ a simple, fast and private (17.) _____ of communication. This combined with the low per-message (18.) _____, the ability to (19.) _____ messages with special characters, emoticons, (20.) _____ and small animations, (21.) _____ to write in English or Japanese, explains (22.) _____ text-messaging is so popular in (23.) _____ among users from every (24.) _____. Teenagers are particularly fond of this form of communication, and many schools ban the use of keitai on (25.) _____.

The inclusion of so many capabilities into a single hand-held device is without doubt a remarkable feat of modern technology. The necessity of these features is another issue, but for many people in Japan a keitai is something they cannot live without.

1. models
2. there
3. can
4. cards
5. most
6. through
7. Japan
8. browse
9. and
10. sites
11. the
12. their
13. a
14. so
15. e-mail
16. is
17. form
18. cost
19. enhance
20. pictures
21. and
22. why
23. Japan
24. generation
25. campus

Appendix D: EW and SEMAC answers (Form B)

EW	SEMAC
1. models	keitai, versions, types
2. there	
3. can	may,
4. cards	
5. Most	Some, Many,
6. through	via, by, with
7. Japan	
8. browse	open, upload, enter, view
9. and	
10. sites	urls, pages, webpages, ones
11. the	
12. their	the, a
13. a	
14. so	therefore, thus, hence
15. e-mail	keitai
16. is	
17. form	means, style, method, way
18. cost	charge, rate
19. enhance	embellish, send, text, make, write, get
20. pictures	photos, images,
21. and	
22. why	
23. Japan	
24. generation	age, nationality
25. campus	