



The Reading Matrix © 2014
Volume 14, Number 1, April 2014

Phrase Cloze: A Better Measure of Reading?

Karim Sadeghi
Urmia University

ABSTRACT

Cloze procedure, a technique not originally intended for measurement purposes, has come to be known as the most widely experimented upon testing tool in English as a Second/Foreign Language (ESL/EFL). Testing experts have indeed used up their energy by constructing and working on various forms of cloze, with the unhappy result that not one person knows what it is that cloze tests measure. Notwithstanding this assertion, cloze tests are still widely used for research purposes and as a component of some high-stakes international tests. This paper intends to formally introduce a new version of cloze, that is, 'phrase-cloze,' and report the findings of a relevant study with Iranian EFL learners on the validity of phrase cloze as a measure of EFL reading comprehension. To test the null-hypotheses of whether phrase cloze is a valid measure of EFL reading comprehension as measured by the First Certificate in English (FCE) and the Certificate of Advanced English (CAE) Reading Papers, 53 candidates (university students majoring in English) filled in an FCE- or a CAE-dependent phrase cloze along with taking the relevant reading tests. The findings suggest that, as far as correlational validation can be trusted (Sadeghi, 2010, 2013), phrase cloze is no better than its traditional counterparts such as classical or standard cloze for testing reading comprehension. Further results and implications are discussed in the paper.

INTRODUCTION

Cloze procedure is now an old name in the literature on language testing. Having been introduced by Taylor (1953) for the first time as a measure of readability, cloze has since been used for a variety of purposes not least of which are for testing language proficiency and reading comprehension in English as a second language. Cloze tests are still used widely in different formats mainly for measuring text comprehension (Sharp, 2009; Schmitt & Sha, 2009; Miller, DeWitt, McCleary & O'Keefe, 2009), proficiency (Keshavarz & Salimi, 2007), collocational knowledge (Stuart & Eve, 2009; Keshavarz & Salimi, 2007), math skills (Pony, Duhon, Lee, & Key, 2010), as a test of plagiarism (Torres & Roig, 2005), as a test of translation (Ito, 2004), as well as a pedagogical tool (Dastjerdi & Talebinezhad, 2006; Lee, 2008). For ease of construction and scoring and because cloze results have shown acceptable correlations with tools testing a variety of abilities, they have so conveniently been used as a cure-all tool for all measurement problems in second/foreign language settings.

Indeed, there has been no shortage of experiments on cloze, and cloze tests are the most widely experimented upon tests in the history of language testing. For this reason, different versions of cloze have been proposed, constructed, and investigated during the last half of the century. What follows next is a brief review of the varieties of cloze procedure experimented upon by other scholars to pave the way for reporting an experiment on ‘phrase cloze.’ Although the labelling of phrase cloze had been suggested in the literature, no research, we are aware of, has been conducted or widely reported on this version of cloze. The aim of the current paper is, therefore, to formally introduce what is known as ‘phrase cloze’ and report the findings of a study with EFL learners taking such a cloze in an effort to uncover whether such cloze tests are in essence different from other varieties of cloze or not and whether they are to be regarded as tests of some abilities. More specifically, this study was conducted to find an answer to the following two research questions:

1. Is there any relationship between ‘phrase-cloze’ and EFL reading comprehension as measured by FCE reading paper?
2. Is there any relationship between ‘phrase-cloze’ and EFL reading comprehension as measured by CAE reading paper?

The above research questions were answered in the form of null-hypotheses, and the related null-hypotheses were tested at the probability levels of 0.01 and 0.05.

VARIETIES OF CLOZE PROCEDURE

The cloze procedure has seen many attempts toward its modification. The original version of cloze in which deletions are made on a random basis or on an every n th order has been termed ‘standard’ cloze, ‘any-word’ cloze (Rankin, 1970; Weaver & Bickley, 1977), ‘natural’ cloze (Ramanauskas, 1972), ‘pseudo-random’ cloze (Alderson, 1979), ‘fixed-ratio’ cloze (Bachman, 1982), ‘classical’ cloze (Klein-Braley, 1983; Klein-Braley & Raatz, 1984), ‘traditional’ cloze, ‘regular’ cloze (Levenston, Nir, & Blum-Kulka, 1984), ‘random’ cloze, ‘mechanical deletion’ cloze (Markham, 1987), ‘completion’ cloze (Hale, et al., 1989), and ‘fixed-interval’ cloze (Spolsky, 2000).

Two kinds of attempt have already been made to alter the original version of the cloze procedure. One kind has been aimed at making cloze more practical in terms of scoring and has led to the introduction of ‘multiple-choice’ cloze as opposed to ‘open-ended’ or ‘free-response’ cloze. The second attempt has been targeted at the construct validity of the original cloze. Such attempts are based on the idea that to test the comprehension of a text on which cloze is based, the random or every n th deletion procedure is not a representative deletion procedure of the words that are important in comprehension. The assumption is that to test particular aspects of comprehension, certain elements should be deleted. These attempts have yielded varieties of cloze called ‘rational’ cloze tests. ‘Rational’ cloze is a name attributed to any kind of cloze test in which deletions are based on a non-mechanical criterion. ‘Rational’ cloze has received support from researchers in the field of reading comprehension rather than readability. Alderson (1983) argues that if the purpose of cloze tests is to measure reading comprehension or language proficiency rather than readability, then the notion of random deletion should be abandoned in favour of rational deletion “based upon a theory of the nature of language and language learning” (p. 211).

According to Markham (1987, p. 304), in rational cloze tests, only content words (such as substantives, verbs, and modifiers) are deleted rather than articles, prepositions or conjunctions. However, one can argue that deleting items of one type only like prepositions can be considered 'rational' because they do not obey the mechanical every *n*th-word deletion procedure. 'Rational' cloze tests as 'professionally developed' tests are in general intended to measure high-order reading processing by fine-tuning the items (Bensoussan, 1990; Storey, 1997). Depending on the purpose of cloze tests, different types of 'rational' cloze tests can be constructed, as Bachman (1985) suggests. More recently, rational cloze procedure has been used not for the purpose of assessment but for integrated instruction of reading, writing, and vocabulary, as well as teacher-student interaction in ESL (Lee, 2008).

The first researcher to experiment with 'rational' cloze test may have been Greene (1965) who used the term 'modified' cloze instead. Comparing a standard version of cloze test with the 'modified' version in which the researcher evaluated each possible item for its effectiveness and left enough redundancy for its completion, Greene (1965) found the two tests similar in terms of test difficulty, but that 'modified' cloze produced more variance, and thus was a more reliable test with better item characteristics whereby very easy or very difficult items were fewer. Comparing the two tests in terms of item-distribution, he found that rationally selected items (content words) in the 'modified' cloze were distributed differently from randomly selected content words in the 'standard' cloze (*ibid.*).

Ozete (1977) introduced a two-choice cloze, called 'modified cloze,' in an attempt to reduce the interruption in the reading process by removing the writing factor. Such a cloze, also called 'reading-input test' by Carver (1978), reduces the burden of the test taker in that s/he only needs to make a choice, forcing a "systematic interaction between the reader and the passage" (p. 566). In an experiment with both easy and difficult 'modified' cloze tests, Ozete (1977) found these tests reliable indicators of the subjects' ability to cope with the difficulty level of the texts. They also discriminated effectively between students at different levels of instruction at lower levels but not at advanced levels.

'Maze' is the name of three-choice cloze used by Pikulski and Pikulski (1977). The two wrong choices are selected such that one is syntactically correct but semantically inappropriate and the other one is both semantically and syntactically incorrect. Thinking of cloze procedure as a teaching device, Brown (1980) suggests a 'selected-deletion' cloze for use to focus on certain aspects of language. Similarly, a cloze type called 'Sel Del Gap test' or 'selected deletion gap-filling test' was suggested by Bensoussan and Mauranen (1988) in which particular cohesive elements in the text are deleted to measure the reader's sensitivity to cohesive links and his/her understanding of macro-level text structures (cited in Bensoussan, 1990).

In a study with non-native speakers of English, Bachman (1982) factor analysed a 'rational' cloze test in which he had deleted syntactic and cohesive items. The results showed that 'rational' cloze was capable of measuring syntactic and discourse relationships in the text. In another study, experimenting with both 'fixed-ratio' cloze and 'rational' cloze made from a single test, Bachman (1985) found that while both tests were of the same reliability and validity indices and discriminated subjects similarly, the former was more difficult.

To test whether context before a cloze item affects performance on cloze or not, Brown (1983) conducted an experiment with EFL students in which he made a cloze test with 50 items and administered it to two groups of subjects. One group sat the test in the normal way; with the other group, however, after each subject gave his/her answer to an item, the correct answer was revealed. Brown (1983) called the first cloze type 'independent-item' cloze and the second type

‘dependent-item’ cloze. His results showed that while there was no difference in reliability between these two cloze types, subjects in the latter cloze significantly outperformed the other group both in exact- and acceptable-scoring procedures. His results were interpreted to mean that cloze items were sensitive to the amount of textual context provided.

Another type of rational cloze, developed by Deyes (1984) and Levenston et al. (1984), is the ‘discourse’ cloze. In ‘discourse’ cloze, deletions are based on elements that mark the relationships between ideas and propositions in the text, such as cohesive ties and discourse markers (Levenston et al., 1984, Coniam, 1993). According to Storey (1997), deletion of such items allows for the identification of good readers who use macro-level text information, as opposed to poor readers who use only local information. Conducting a study using ‘discourse’ cloze with Hebrew-speaking children, Levenston et al. (1984) found that ‘discourse’ cloze was able “to distinguish between different levels of achievement in reading at least as well as the regular cloze” (p. 210).

Another ‘rational’ cloze developed by Bensoussan and Ramraz (1984) over a period of 10 years is ‘fill-in’ cloze. Contrary to what the name suggests, the test is not in an open-ended format but in four-choice items. The only significant difference between this cloze type and other similar ones is that in ‘fill-in’ cloze there is a possibility of deleting whole phrases or expressions as well as single words. The test is intended to test EFL learners’ reading comprehension.

In an attempt to produce tests of higher reliability and validity using traditional item-analysis procedures, Brown (1988) introduced what he called ‘tailored cloze.’ In his study, Brown piloted a text of 399 words with 50 deletions to a group of 89 EFL learners. First time, all students received a 7th deletion rate cloze. After 6 weeks, four other versions of the same text with the same deletion rate but different starting points were administered to the same subjects in groups of 22 to 23. Finally, he selected 50 ‘best’ items out of a total of 250 items, based on item facility and item discrimination indices. The cloze test was reconstructed deleting these 50 items and was called ‘tailored cloze.’ The item-analysis of cloze items originated with Greene (1965), however, and was followed by Cranney (1972-1973) who constructed two tailored cloze tests: one for a free-response and another for an m/c cloze. In free-response tailored cloze both reliability and validity dropped compared to the original 300-item cloze. For multiple-choice tailored cloze, however, both reliability and validity improved. Cranney (1972-73) concluded that m/c tailored cloze was a better test than free-response tailored cloze.

‘Semantic’ cloze is an m/c version of cloze developed by Mauranen (1989) to avoid productive skills and make scoring as objective and quick as possible. In ‘semantic’ cloze, which is intended to test EFL reading comprehension, the focus is on “direct measurement of the comprehension of meaning and higher range (macro level) items” (p. 341). It shares a lot with other similar tests developed by Bachman (1982), Levenston et al. (1984), and Bensoussan and Ramraz (1984), but it is different from them in that more advanced level academic texts are used in test-construction and is intended to test only *textual comprehension* of sentence-level or above, but not extra-textual or pragmatic knowledge. Comparing ‘semantic’ cloze with ‘standard’ cloze, Mauranen (1989, p. 342) points out that the former is more text-sensitive than the latter but is as difficult to construct as a multiple-choice ‘standard’ cloze test. The ‘semantic’ cloze was suggested to accompany other reading tests in a test battery to allow for the measurement of what cloze cannot test, that is, “reader’s independent and critical interpretation of the main points of a text” (p. 343).

Aimed at testing non-native English speakers’ understanding of text redundancy and cohesion, ‘cohesion’ cloze was introduced by Bensoussan (1990). In ‘cohesion’ cloze, subjects

are required to fill in the blanks using the words in the text, and blanks appear wherever redundancy allows, which means that there can be a single word or a whole paragraph between two blanks (p. 26). In constructing 'cohesion' cloze, the third and later occurrences of a word or its cohesive link are deleted. Based on her study with 'cohesion' cloze, Bensoussan concluded that while 'cohesion' cloze may not be suitable for readability or language proficiency purposes, it is a means of evaluating the reader's micro- and macro-level comprehension processes, and that it can be used as a self-teaching device in training to identify contextual clues.

'Summary' cloze is another version of rational-deletion cloze widely used in Hong Kong (Coniam, 1993; Storey, 1997). 'Summary' cloze, which was once a part of the International English Language Testing System (IELTS), is based on a text which students read in full before doing the cloze. The cloze test itself is supposed to be a summary of the original text, and its completion is expected to show the degree of comprehension of the original text. 'Summary' cloze, also called 'paraphrasing' cloze, is intended to test comprehension indirectly, objectively, and at discourse level (Coniam, 1993). Based on his experiment with a 'summary' cloze, which was validated against an m/c 'standard' cloze, he found that the presence of the original text before the 'summary' cloze did not affect test results in terms of its validity. Coniam (1993) concluded that although 'summary' cloze tests some element of comprehension and summarizing, it also taps something more general not related to the preceding passage. Based on such a conclusion, he regards 'summary' cloze as an invalid test of reading but as a valid and viable test of general language proficiency (p. 9). According to Coniam (1993), 'summary' cloze was not intended originally to be a version of cloze, but to be a test of reading comprehension called 'summary completion' (ibid.).

'Oral' cloze is the name for the spoken version of cloze procedure. In 'oral' cloze, the subject listens either to a tape in which a passage is read out in cloze format with pauses when actual blanks occur, or to a person reading the text live. In an experiment with 'oral' cloze, Taylor (1956) found such tests to discriminate between the 'listenabilities' of spoken discourse. Whether the same mental load is placed on subjects in oral cloze as in written cloze is not settled (Davies, 1979). To allow for the extra mental burden, Weaver (1977) suggests an alternative, which he calls 'multiple-pass situation.' Namely, subjects can hear the cloze as many times as they require. Comparing written cloze performance with one-pass and multiple-pass 'oral' cloze performance, Weaver (1977) found that while 'structural meaning' was understood better in written cloze, 'lexical meaning' or the prediction of nouns and verbs is equally understood in both types (ibid.). 'Oral' cloze tests have also been called 'Auditory' cloze tests. Ulusoy (2010) used such ACT's (auditory cloze tests) to determine the listening ability level of students from Ankara while listening to Internet Radio programs

Considering cloze an invalid measure of EFL language proficiency, Klein-Braley and Raatz (1984) introduced a similar method in 1981 called 'C-test' based on 'the rule of 2.' In these tests, the second part of every second word is deleted. Klein-Braley and Raatz (Klein-Braley, 1983, 1985; and Klein-Braley & Raatz, 1984) were two main scholars who argued for and conducted numerous experiments in favour of C-tests. They contend that 'C-tests' are valid and reliable measures of EFL language proficiency. Similarly, working with Hungarian EFL learners, Dörnyei and Katona (1992) found 'C-tests' to be reliable and valid as a measure of general language proficiency. Chapelle and Abraham (1990), however, question the validity of 'C-tests' and claim that although they seem to measure grammatical competence more than lexical ability, it is not clear at all what such tests measure. Although Klein-Braley and Raatz (1984) claim C-tests to be both empirically and theoretically valid, and that native speakers can

obtain scores of 100% on them, Jafarpur (1996) did not find them to be valid measures of EFL language proficiency, because only in rare cases did the native subjects approach the perfect score in his experiment. Although his study was criticized by Hastings (2002a, 2002b) for design problems, Jafarpur's later experiments showed the superiority of C-tests over 'standard' cloze tests in terms of reliability and concurrent validity (Jafarpur, 2002a, 2002b). Pino and Eksenazi (2009) experimented with a similar sentence-level open cloze in which they provided the candidates with blanked sentences where the first few letters were given as a hint to help candidates restore the missing word.

A rather different kind of cloze, which may not even be considered cloze, was designed and studied by Manning (1987). 'Cloze-elide' tests are constructed not by deleting words from the text but by inserting extraneous words at random positions in the text and asking the subjects to detect them. Manning (1987) used 'cloze-elide' tests along with Test of English as a Foreign Language (TOEFL), an essay, and a 'random' cloze, and found them to enjoy strong concurrent validity. Factor analysis showed that all tests were 'good, indirect measures of English language proficiency,' and were also able to predict teachers' ratings of students' proficiencies.

Other tasks have been reported which may be similar to cloze in form but are different in purpose. 'Sentence-completion' tasks are similar to cloze in that there are blanks to be completed as there are in cloze tests. The differences have been elaborated on by Taylor (1953) and Rankin (1970). While the former tests are constructed to measure one's discrete-point knowledge of certain items, and the sentences may be in no way connected to each other, in 'standard' cloze tests, blanks are not intentionally selected, and the whole sequence of sentences forms a coherent piece of discourse. As Rankin (1970) states, answering an item rightly or wrongly has no influence on correct answering of the following items in 'sentence-completion' tasks, but it may affect the predictability of following items in cloze tests. While there have been suggestions in the literature as to the possibility of deleting phrases rather than single words as noted above, we have been unable to track any studies on phrase-level cloze.

METHOD

Participants

A total of 53 participants took part in this study. They were all university students majoring in English Language and Literature at Urmia University, Iran. 34 students took the FCE Reading Paper as well as the FCE-related phrase cloze, and 19 candidates took the CAE Reading Paper along with the relevant phrase-cloze. Most FCE takers were freshmen and sophomores (relatively lower proficient group) and most CAE takers were juniors and seniors (relatively higher proficient group). The majority of the candidates spoke Azeri as their L1, and all spoke Persian fluently (for some it was their first language). Few candidates spoke Kurdish as their L1. Table 1 shows the characteristics of the participants in the study.

Table 1. Characteristics of the Participants in the Study

Characteristics Tests	n	Sex			Age	
		M	F	Unknown	Mean	Range
FCE Reading Paper/FCE-based phrase cloze	34	7	26	1	20	16
CAE Reading Paper/CAE-based phrase cloze	19	6	13	0	22.95	11

Instruments

The major instrument used in the study was a phrase cloze-test. When quantitative studies of validation are conducted, there is a need to use another criterion test, which is supposed to be valid *a priori*. Considering that the nature of reading comprehension is yet far from being completely understood, it can be asserted that there may be no valid test of reading comprehension at the present moment. Although it is accepted here that with the present state of our knowledge about reading and testing, we are way beyond having a valid test of reading comprehension, it is, however, acknowledged that one test may be better than another for this purpose. With this in mind, a review of the proficiency tests made by Cambridge ESOL was undertaken, and because they are said to undergo a constant process of validation whereby all EFL tests are trailed, piloted, analysed, and reviewed before they are made into final test papers (Saville, 2001; UCLES, 2000, 2001), they were found to be better than their counterparts produced by ETS and other test-making bodies. As such, among the tests reviewed, the FCE, the CPE, the CAE, and the IELTS were regarded as suitable as criterion measures.

A pilot study with IELTS proved it a very difficult test for the participants in question. In addition, as the CPE is intended to measure the language ability of EFL teachers, it was regarded to be more advanced for the level of the subjects who were going to participate in this study. Therefore, the only options available were the FCE (First Certificate in English) and the CAE (Certificate in Advanced English), both of which were used in the study. The former is easier than the latter and suitable for the first year and low-level university students majoring in English; the latter is relatively more difficult and more suitable for more advanced subjects as those in year 3 and 4 of the university.

To avoid the problem of finding a passage for cloze-test which would have the same difficulty level as the criterion test, the researcher was certain that regardless of method used, no text could be found with the same difficulty level (not only in terms of lexical and grammatical meaning but also in terms of semantics and pragmatics) as the text in the criterion test. The best solution to this problem was to use the same passage used for the criterion reading test such as the cloze passage. Herein, the cloze passage would be exactly the same passage used in the reading test and the results would be conveniently comparable. Of course, if the reading test were to be taken before the cloze test, there would be a strong carry-over from the reading test to the cloze test. However, this potential danger was avoided in this study by administering the cloze tests prior to the reading tests.

Although the possibility of 'phrase-cloze' has been suggested earlier, there has been no research on such cloze tests as far as this researcher is aware. To construct 'phrase-cloze,' phrases between 2 and 4 words in length were left out from each passage, and the candidates were instructed to fill in the blanks with 'phrases' rather than single words. Because the participants were unfamiliar with this type of cloze, a few had filled in the blanks with words rather than phrases. Both FCE- and CAE-based 'phrase-cloze' tests had 40 items. (A copy of the relevant cloze tests appear in the appendix.)

Procedure

During the data collection, the purpose of the research was explained to students in each class orally as well as appearing on the covering letter given to each student with the test materials. Each subject also received an answer sheet in which they were asked to give their name (optional), student number, age, sex, languages spoken except for Farsi and English, and

whether they attended private English classes.

The relevant test materials were administered to each group in different sessions and during their normal class hours. In each session, the covering letter and the answer sheet were given first, and the candidates were given time to read and write the relevant information. Then, the phrase cloze tests were administered first. The time given for the cloze test on the paper was 25 minutes, but it was increased to half an hour to obtain enough data from the majority of students. Instructions were also given to the students to write all the answers on the answer sheet. After the time for the cloze test (Part 1 of the test) was over, the cloze papers and answer sheets were collected and Part 2 or the FCE/CAE reading test was administered. The reading test was given 20 minutes, and the time seemed to be sufficient for most students.

The cloze tests were scored for both exact scoring and acceptable scoring. In acceptable scoring two types of acceptable-scoring were used. The first time, all acceptable ‘phrases’ only, and the second time, all acceptable ‘phrases’ and acceptable ‘words’ were given credit.

RESULTS

Tables 2 and 3 represent statistics for the test sets of the FCE and the CAE respectively in terms of the number of participants who took each test, mean, range, and standard deviation (SD). For the ease of comparison, all means have been converted to percentages.

Table 2. Descriptive Characteristics for the FCE Test Set with Both Exact- and Acceptable-scoring for Phrase Cloze Test

Characteristics Tests	N	mean (%)	range	SD
Exact	34	0.8	4	0.77
Acceptable word + phrase	34	18.37	25	6.05
Acceptable phrase	34	12.73	24	5.38
TRC	34	31.75	6	1.64

Table 3. Descriptive Characteristics for the CAE Test Set with Both Exact- and Acceptable-scoring for Phrase Cloze Test

Characteristics Tests	N	mean (%)	range	SD
Exact	19	1.7	3	0.95
Acceptable word + phrase	19	24.72	21	5.56
Acceptable phrase	19	13.15	14	3.87
TRC	19	33.33	10	2.33

To answer the two research questions posed in the Introduction, data were used from Tables 4 and 5. In both cases, not only was the relationship between ‘phrase-cloze’ and TRC explored for exact-scoring of cloze but for two kinds of acceptable-scoring, that is, acceptable ‘phrase’ scoring, in which only acceptable ‘phrases’ were given credit, and acceptable ‘word + phrase’ scoring, in which acceptable ‘words’ were also counted correct in addition to acceptable ‘phrases.’ As Tables 4-5 indicate, not only was there no significant relationship between ‘phrase-cloze’ and TRC in either the FCE or the CAE, but the relationships between ‘phrase-cloze’ tests and TRC’s were actually negative in some cases.

Table 4. Correlation Coefficients between the FCE ‘Phrase-cloze’ Test and TRC

Tests	exact-scoring	acceptable-word + phrase scoring	acceptable-phrase scoring	TRC
Exact-scoring	1.00	—	—	—
Acceptable-word + phrase scoring		1.00	0.815*	-0.122
Acceptable-phrase scoring			1.00	.005
TRC		1.00		

Note. Significant at 0.01= *

Table 5. Correlation Coefficients between the CAE ‘Phrase-cloze’ Test and TRC

Tests	exact-scoring	acceptable-word + phrase scoring	acceptable-phrase scoring	TRC
Exact-scoring	1.00	—	—	—
Acceptable-word + phrase scoring		1.00	0.877*	0.206
Acceptable-phrase scoring			1.00	.166
TRC		1.00		

Note. Significant at 0.01= *

It is worth noting that a correlation was not run between exact-scoring of ‘phrase-cloze’ tests and the relevant TRC’s neither for the FCE nor for the CAE. Neither was a correlation run between the exact-scoring and acceptable ‘phrase’ and acceptable ‘word + phrase’ scoring of the ‘phrase-cloze,’ because, as Tables 2 and 3 indicate, the amount of data produced in exact-scoring of these cloze tests was so small, that no meaningful variation could be produced as a result. With the added problem of the fewer number of subjects in the case of the CAE, the relevant validity indices could be even less meaningful if calculated. As far as the relationship between ‘phrase-cloze’ and TRC is concerned, the positive relationship in both cases (FCE and CAE) is rather weak. The fact that none of the positive and negative relationships in either the FCE or the CAE is significant in any of the scoring methods invites us to safely confirm the related null-hypotheses. Thus, the two research questions posed in this study can be answered in the following way:

1. There is no statistically significant relationship between ‘phrase-cloze’ test and EFL reading comprehension as measured by FCE reading test.
2. There is no statistically significant relationship between ‘phrase-cloze’ test and EFL reading comprehension as measured by CAE reading test.

DISCUSSION AND CONCLUSION

No statistically significant relationship was observed between ‘phrase-cloze’ tests and related reading comprehension tests in both the FCE and the CAE. As such, the evidence produced above suggests that phrase-cloze tests were not significantly related to corresponding reading tests. Such a finding is usually statistically interpreted as an indication that cloze tests (as

employed in this study) are not valid measures of reading comprehension (as measured by TRC's in this study). For the purpose of this study, however, it means that 'phrase' cloze tests are not valid measures of EFL reading comprehension in both the FCE and the CAE. It is worth noting here that such a research trend in language testing in which tests are validated simply based on a degree of correlation has already been challenged (Sadeghi, 2002, 2006, 2010, 2013).

It has been argued that while it is acceptable that two variables which replace each other should correlate to a high degree, the reverse side of the issue (i.e., because two measures are highly correlated, they can replace each other) is challenged. It has also been argued that a test may be suggested to replace another and still be considered a valid measure of that test if all the following conditions are met:

- a) If both tests are of the same nature (cloze tests, for example);
- b) If both tests are intended for the same purpose (for testing reading comprehension, for example); and
- c) If the correlation between the two tests is not only significant but also very high and near +1.00; and if one wishes to lose no information by replacing one with another, their correlation and shared variance should be perfect, something which may be possible only in theory.

The fact that cloze tests and reading tests are not of the same nature or character (which violates the first condition above) prevents us from concluding that 'phrase-cloze' tests are valid measures of reading comprehension even if the resulting correlation coefficients were near perfect. From the preceding argument then, it can be surmised that even in cases where there are statistically significant relationships between cloze tests and reading comprehension tests, concluding that cloze tests are valid measures of reading comprehension may not be justifiable.

This paper investigated for the first time the problem of whether 'phrase-cloze' is a valid measure of EFL reading comprehension in quantitative terms. The statistical tests used showed that the phrase cloze tests used here were not significantly related to the reading comprehension tests. Accepting that different cloze tests pose different requirements on the reader mainly because of the nature of the texts used and the type of deletions made, it was argued that concluding that cloze tests are valid measures of EFL reading comprehension simply because of a correlation figure may not be justifiable. The issue of what phrase cloze tests measure and their validity as a measure of EFL reading comprehension still remains open for further research. In closing, it is proposed here that the problem at hand may not be well addressed using quantitative tools only and that a mixture of quantitative and qualitative research procedures may reveal a much better picture of the reality in question.

Karim Sadeghi holds a PhD in TESOL from the University of East Anglia (UK) and is an academic member of staff in Urmia University, Iran. His recent research has focused on language assessment as well as research evaluation. He has served as an editorial/advisory board member of numerous international and national journals and is currently the editor-in-chief of the Iranian Journal of Language Teaching Research. In 2013, he was selected as Iran's top researcher in Humanities and Social Sciences.

Email: k.sadeghi@urmia.ac.ir

REFERENCES

- Alderson, J. C. (1983). The cloze procedure and proficiency in English as a foreign language. In J. W. Oller (ed.), *Issues in language testing research* (pp. 205-212). Rowley, MA: Newbury House Publishers.
- Alderson, J. C. (1979). The cloze procedure and proficiency in English as a foreign language. *TESOL Quarterly*, 13, 219-228.
- Bachman, L. F. (1985). Performance on the cloze test with fixed ratio and rational deletions. *TESOL Quarterly*, 19, 335-356.
- Bachman, L. F. (1982). The trait structure of cloze test scores. *TESOL Quarterly*, 16, 61-70.
- Bensoussan, M. (1990). Redundancy and the cohesion cloze. *Journal of Research in Reading*, 13, 18-37.
- Bensoussan, M., & Ramraz, R. (1984). Testing EFL reading comprehension using a multiple-choice rational cloze. *Modern Language Journal*, 68, 230-239.
- Brown, J. D. (1988). Tailored cloze: Improved with classical item analysis techniques. *Language Testing*, 5, 19-31.
- Brown, J. D. (1983). A closer look at cloze: Validity and reliability. In J. W. Oller (Ed.), *Issues in language testing research* (pp. 237-250). Rowley, MA: Newbury House Publishers.
- Brown, J. D. (1980). Relative merits of four methods for scoring cloze tests. *Modern Language Journal*, 64, 311-317.
- Carver, D. (1978). Reading comprehension – Is there such a thing? *ELT Journal*, 32, 291-297.
- Chapelle, C. A., & Abraham, R. G. (1990). Cloze method: What difference does it make? *Language Testing*, 7, 121-146.
- Coniam, D. (1993). Co-text or no text? A study of an adapted cloze technique for testing comprehension/summary skills. *Hong Kong Papers in Linguistics and Language Teaching*, 16, 1-10.
- Cranney, A. G. (1972-1973). The construction of two types of reading tests for college students. *Journal of Reading Behavior*, 5, 60-64.
- Dastjerdi, H. V., & Talebinezhad, M. R. (2006). Chain-processing deletion procedure in cloze: A discourse perspective. *Language Testing*, 23, 58-72.
- Davies, P. (1979). Cloze tests. In M. J. Raggett, C. Tutt, & P. Raggett (Eds.), *Assessment and testing of reading: Problems and practices* (pp. 62-69). London: Ward Lock Educational.
- Deyes, T. (1984). Towards an authentic 'Discourse Cloze'. *Applied Linguistics*, 5, 128-137.
- Dörnyei, Z., & Katona, L. (1992). Validation of the C-test amongst Hungarian EFL learners. *Language Testing*, 9, 187-206.
- Greene, F. P. (1965). Modifications of the cloze procedure and changes in reading tests performance. *Journal of Educational Measurement*, 2, 213-217.
- Hale, G. A., Stansfield, C. W., Rock, D. A., Hicks, M. M., Butter, F. A., & Oller, J. W. (1989). The relation of multiple-choice cloze items to the Test of English as a Foreign Language. *Language Testing*, 6, 47-76.
- Hastings, A. J. (2002a). In defence of C-testing. In R. Grotjahn, (Ed.), *Der C-Test: Theoretische Grundlagen und praktische Anwendungen (Band 4)* (pp. 11-25). Bochum: AKS-Verlag Bockum.
- Hastings, A. J. (2002b). A response to Jafarpur. In R. Grotjahn, (Ed.), *Der C-Test: Theoretische Grundlagen und praktische Anwendungen (Band 4)* (pp. 28-29). Bochum: AKS-Verlag Bockum.

- Ito, A. (2004). Two types of translation tests: Their reliability and validity. *System*, 32, 395-405.
- Jafarpur, A. (2002a). A response to Hastings. In R. Grotjahn (Ed.), *Der C-Test: Theoretische Grundlagen und praktische Anwendungen (Band 4)* (pp. 26-27). Bochum: AKS-Verlag Bockum.
- Jafarpur, A. (2002b). *A comparative study of a C-test and a cloze test*. In R. Grotjahn (Ed.), *Der C-Test: Theoretische Grundlagen und praktische Anwendungen (Band 4)* (pp. 31-51). Bochum: AKS-Verlag Bockum.
- Jafarpur, A. (1996). Native speaker performance validity: In vain or for gain? *System*, 24, 83-95.
- Keshavarz, M. H., & Salimi, H. (2007). Collocational competence and cloze test performance: A study of Iranian EFL learners. *International Journal of Applied Linguistics*, 17, 81-92.
- Klein-Braley, C. (1983). A cloze is a cloze is a question. In J. W. Oller (Ed.), *Issues in language testing research* (pp. 218-228). Rowley, MA: Newbury House Publishers.
- Klein-Braley, C. (1985). A cloze-up on the C-test: A study in the construct validation of authentic tests. *Language Testing*, 2, 76-104.
- Klein-Braley, C., & Raatz, U. (1984). A survey of research on the C-test. *Language Testing*, 1, 134-146.
- Lee, S. H. (2008). Beyond reading and proficiency assessment: The rational cloze procedure as stimulus for integrated reading, writing, and vocabulary instruction and teacher-student interaction in ESL. *System*, 36, 642-660.
- Levenston, E. A., Nir, R., & Blum-Kulka, S. (1984). Discourse analysis and the testing of reading comprehension by cloze techniques. In A. K. Pugh, & J. M. Ulijn (Eds.), *Reading for professional purposes* (pp. 202-212). London: Heinemann Educational Books.
- Manning, W. H. (1987). *Development of cloze-elide tests of English as a second language (TOEFL Research Report 23)*. Princeton, NJ: Educational Testing Service.
- Markham, P. L. (1987). Rational deletion cloze processing strategies: ESL and native English. *System*, 15, 303-11.
- Mauranen, A. (1989). Can gaps measure comprehension? Modifications of cloze as tests of reading. In C. Lauren & M. Nordman (Eds.), *Special language: From humans thinking to thinking machines* (pp. 337-346). Clevedon & Philadelphia: Multilingual Matters.
- Miller, M. J., DeWitt, J. E., McCleary, E. M., & O'Keefe, K. J. (2009). Application of the cloze procedure to evaluate comprehension and demonstrate rewriting of pharmacy educational materials. *Annals of Pharmacotherapy*, 43, 650-657.
- Ozete, O. (1977). The cloze procedure: A modification. *Foreign Language Annals*, 10, 565-568.
- Pikulski, J. J., & Pikulski, E. C. (1977). Cloze, maze, and teacher judgement. *The Reading Teacher*, 30, 766-770.
- Pino, J., & Eskenazi, M. (2009). Measuring hint level in open cloze questions. *Proceedings of the 22nd International Florida Artificial Intelligence Research Society Conference, FLAIRS-22* (pp. 460-465).
- Pony, B. C., Duhon, G. J., Lee, S. B., & Key, A. (2010). Evaluation of techniques to promote generalization with basic math fact skills. *Journal of Behavioral Sciences*, 19, 76-92.
- Ramanauskas, S. (1972). The responsiveness of cloze readability measures to linguistic. Variables operating over segments of text longer than a sentence. *Reading Research Quarterly*, 8, 72-91.
- Rankin, E. F. (1970). The cloze procedure – Its validity and utility. In R. Farr (Ed.), *Measurement and evaluation of reading* (pp. 237-253). New York, NY: Harcourt, Brace and World.
- Sadeghi, K. (2013). Doubts on the validity of correlation as a validation tool in second language

- testing research: The case of cloze testing. *Language Testing in Asia Journal*, 3(15), 1-17.
- Sadeghi, K. (2010). Cloze validation against IELTS Reading Paper: Doubts on correlational validation. *Journal of English Language Teaching and Learning*, 53(2) 115-139.
- Sadeghi, K. (2006). *Rethinking correlational validation*. Paper presented at the 3rd TELLSI Conference, Razi University, Kermanshah, Iran.
- Sadeghi, K. (2002). *Is correlation a valid statistical tool in second language research?* Paper presented at the 12th European Second Language Association (EUROSLA12) Conference, University of Basel, Basel, Switzerland.
- Saville, N. (2001). *Test development and revision*. In UCLES. *Research Notes No. 4* (pp. 5-8). Cambridge: UCLES.
- Schmitt, M. C., & Sha, S. (2009). The developmental nature of meta-cognition and the relationship between knowledge and control over time. *Journal of Research in Reading*, 32, 254-271.
- Sharp, A. (2009). Reading comprehension in two cultures. *International Journal of Learning*, 16, 281-292.
- Spolsky, B. (2000). Language testing in the Modern Language Journal. *Modern Language Journal*, 84, 536-552.
- Storey, P. (1997). Examining the test-taking process: A cognitive perspective on the discourse cloze test. *Language Testing*, 14, 214-231.
- Stuart, W., & Eve, K. (2009). The effects of vocabulary learning on collocation and meaning. *TESOL Quarterly*, 43, 55-77.
- Taylor, W. L. (1956). Recent developments in the use of 'cloze procedure.' *Journalism Quarterly*, 33, 42-49 & 99.
- Taylor, W. L. (1953). Cloze procedure – A new tool for measuring readability. *Journalism Quarterly*, 30, 415-433.
- Torres, M., & Roig, M. (2005). The cloze procedure as a test of plagiarism: The influence of text readability. *Journal of Psychology: Interdisciplinary and Applied*, 139, 221-231.
- UCLES (2001). *FCE handbook*. Cambridge: UCLES.
- UCLES (2000). *IELTS handbook*. Cambridge: UCLES.
- Ulusoy, M. (2010). Assessing the listening ability levels of Internet radio programs. *New Education Review*, 20, 306-324.
- Weaver, W. W. (1977). An examination of some differences in oral and written language using the cloze procedure. In A. J. Kingston (Ed.), *Toward a psychology of reading and language: Selected writings of Wendell W. Weaver* (pp. 96-97). Athens, GA: University of Georgia Press.
- Weaver, W. W., & Bickley, A. C. (1977). An empirical examination of cloze scores derived from 'natural' and 'mutilated' language segments. In A. J. Kingston (Ed.) *Toward a psychology of reading and language: Selected writings of Wendell W. Weaver* (pp. 140-149). Athens, GA: University of Georgia Press.

APPENDIX

FCE Phrase-cloze

On Saturday mornings I worked in ...1... . I started cycling down to the shop with Dad ...2... as soon as I was big enough. I thought of it as ...3... and so I didn't mind what I did, although it was mostly just fetching and carrying at a run all morning. I managed not to think of it ...4... and I looked forward to the bar of chocolate my grandmother ...5... unsmilingly as I left. I tried not to look at her; I had reason to feel guilty because I'd generally already eaten some dried fruits or a silver of cheese when no one was looking. As soon as ...6..., though, Dad said, 'That's it, our Janet. You're of working age now and you are not ...7... unless your grandmother pays you properly.' He did his best to make his chin look determined. 'I shall speak to her.'

The next Saturday, Gran called me into ...8... behind the shop. I always hated ...9... there. She had an electric heater on full blast, and the windows were always kept tightly closed ...10... . 'You're wanting to ...11..., I hear,' she said. 'Yes, please,' I replied. I was rather like visiting the headmistress ...12..., so I was very quiet and respectful. Gran searched through the mess of papers on ...13..., sighing and clicking her tongue. Eventually she produced ...14... and ran her fingers along the columns of figures. '...15...' 'Fifteen ... Gran,' I added for extra politeness, but she looked at me as if I had been cheeky. 'Full-timers ...16... get forty pounds for a thirty-five-hour week,' she announced in ...17... as to leave no doubt that she wasn't in favour of this. 'No wonder there's no profit ...18...! So, Janet, what's that per hour?' Questions like that always flustered me. Instead of trying to ...19... in my head, I would just stand there, unable to think straight. 'I'll get ...20..., ' I offered. 'Don't bother,' snapped Gran angrily, 'I'll do it myself. I'll give you a pound an hour; take it or ...21... . 'I'll take it, please.' 'And I expect real work for it, mind. No standing about, and if I catch you eating any of the stock, there will be trouble. ...22..., and it's a crime.'

From then on, my job ...23... was filling the shelves. This was dull, but I hardly expected to be trusted with ...24... . Once or twice, however, when Dad was extra busy, I'd tried to help him by serving ...25... . I hated it. It was very difficult to remember the prices of everything and I was particularly hopeless at using the till. ...26... made unkind remarks about this, increasing my confusion and the chances of my making a ...27... .

It was an old-established village shop, going back 150 years at least and it was really ...28... even then. Dad longed to be able to ...29... more attractive to customers, but Gran wouldn't hear of it. I overheard them once arguing about whether to buy a freezer cabinet. '...30... want frozen food,' Dad said. 'They see things advertised and if they can't get them from us, they will ...31... .' 'Your father always sold fresh food,' ...32... . 'People come here for quality, they don't want all that ...33... .'

Actually, she gave way ...34... over the freezer. Mr Timson, her great rival, installed one ...35... at the other end of the village and customers started making loud comments about how handy it was, being able to get ...36... in the village, and how good ...37... sausages were. That really upset her because she was proud of ...38... and she ungraciously gave Dad the money to ...39... . Within a couple of weeks, ...40... frozen food like the rest of us.

FCE Reading Comprehension Test

On Saturday mornings I worked in the family shop. I started cycling down to the shop with Dad on Saturdays as soon as I was big enough. I thought of it as giving him a hand and so I didn't mind what I did, although it was mostly just fetching and carrying at a run all morning. I managed not to think of it as work and I looked forward to the bar of chocolate my grandmother passed me unsmilingly as I left. I tried not to look at her; I had reason to feel guilty because I'd generally already eaten some dried fruits or a silver of cheese when no one was looking. As soon as I was fifteen, though, Dad said, 'That's it, our Janet. You're of working age now and you are not coming to work unless your grandmother pays you properly.' He did his best to make his chin look determined. 'I shall speak to her.'

The next Saturday, Gran called me into her little office behind the shop. I always hated going in there. She had an electric heater on full blast, and the windows were always kept tightly closed whatever the weather. 'You're wanting to get paid, I hear,' she said. 'Yes, please,' I replied. I was rather like visiting the headmistress at school, so I was very quiet and respectful. Gran searched through the mess of papers on her crowded desk, sighing and clicking her tongue. Eventually she produced an official-looking leaflet and ran her fingers along the columns of figures. 'How old are you?' 'Fifteen ... Gran,' I added for extra politeness, but she looked at me as if I had been cheeky. 'Full-timers at your age get forty pounds for a thirty-five-hour week,' she announced in such a way as to leave no doubt that she wasn't in favour of this. 'No wonder there's no profit in shopkeeping! So, Janet, what's that per hour?' Questions like that always flustered me. Instead of trying to work them out in my head, I would just stand there, unable to think straight. 'I'll get a pencil and paper,' I offered. 'Don't bother,' snapped Gran angrily, 'I'll do it myself. I'll give you a pound an hour; take it or leave it.' 'I'll take it, please.' 'And I expect real work for it, mind. No standing about, and if I catch you eating any of the stock, there will be trouble. That's theft, and it's a crime.'

From then on, my job at the shop was filling the shelves. This was dull, but I hardly expected to be trusted with handling the money. Once or twice, however, when Dad was extra busy, I'd tried to help him by serving behind the counter. I hated it. It was very difficult to remember the prices of everything and I was particularly hopeless at using the till. Certain customers made unkind remarks about this, increasing my confusion and the chances of my making a fool of myself.

It was an old-established village shop, going back 150 years at least and it was really behind the times even then. Dad longed to be able to make the shop more attractive to customers, but Gran wouldn't hear of it. I overheard them once arguing about whether to buy a freezer cabinet. 'Our customers want frozen food,' Dad said. 'They see things advertised and if they can't get them from us, they will go elsewhere.' 'Your father always sold fresh food,' Gran replied. 'People come here for quality, they don't want all that frozen stuff.'

Actually, she gave way in the end over the freezer. Mr Timson, her great rival, installed one in his shop at the other end of the village and customers started making loud comments about how handy it was, being able to get frozen food in the village, and how good Mr Timson's sausages were. That really upset her because she was proud of her sausages and she ungraciously gave Dad the money to buy the freezer. Within a couple of weeks, she was eating frozen food like the rest of us.

- 1) How did Janet feel when she first started her Saturday morning job?
 - A. She enjoyed the work that she was given.
 - 1) She was pleased to be helping her father.
 - 2) She worried that she was not doing it well.
 - 3) She was only really interested in the reward.
- 2) What do we learn about her grandmother's office in paragraph two?
 - 1) It needed decorating.
 - 2) It was untidy.
 - 3) It had too much furniture in it.
 - 4) It was dark.
- 3) 'This' (line 25, underlined) refers to
 2. shopkeepers' profits.
 3. a thirty-five-hour week.
 4. Janet's request.
 5. the recommended wage.
- 4) 'Flustered' (line 26, underlined) means
 1. bored.
 2. angered.
 3. confused.
 4. depressed.
- 5) Why did Janet's grandmother react angrily to her offer to fetch a pencil and paper?
 - a) Janet was unable to answer her question.
 - b) Janet had been unwilling to help her.
 - c) Janet had made an unhelpful suggestion.
 - d) Janet had answered her rudely.
- 6) What did Janet's father and grandmother disagree about?
 - A. how to keep their customers loyal to the shop.
 - B. the type of advertising needed to attract customers.
 - C. the type of customers they needed to attract.
 - D. how to get new customers to come to the shop.
7. What eventually persuaded Janet's grandmother to buy a freezer?
 - A. She found that she liked frozen food after all.
 - B. A new shop opening in the village had one.
 - C. It was suggested that her products weren't fresh.
 - D. She responded to pressure from her customers.
8. What impression do we get of Janet's feelings towards her grandmother?
 - A. She respected her fairness.
 - B. She doubted her judgement.
 - C. She disliked her manner.
 - D. She admired her determination.

CAE Phrase-cloze

Labels

The paper label must be one of the most abundant of art forms and, ...1... the number of people it reaches, ...2... influential. Here is a brief look at the history and functions of the labels on everyday products.

Manufacturers have become increasingly aware of ...3... a part the label plays in the sales of ...4.... Considerable attention is now given to the visual impact ...5..., particularly its overall colour, this being the first point of ...6.... In the early days, brands were sometimes asked ...7... only, to such an extent that the company would incorporate the colour into ...9.... Thus, there was 'Green Label' chutney and 'Red Label' tea.

Manufacturers ought to remember that the label can become so opulent in comparison with ...10... that the customer's expectations are dashed when the contents are revealed. In 1888, when ...11... in the US was moving towards ever more lavish labels, an article in the New York Sun commented, 'The label is often ...12... the cigar...'

Since label design plays such an important role in a product's continuity, it is surprising that manufacturers today allow labels to be vandalised by a clutter of competition announcements and ...13.... One can only suppose that the public are not offended by these recent design changes, which are now commonplace ...14... – 'the gallery of commercial art' – where some 2,000 designs are displayed. The variety is certainly incredible. There are those ...15... follow the latest fashion, those which are modern but not 'over the top', those which retain much of ...16... previous generations yet are still smart, and ...17... try to look nostalgic. Design apart, the label itself has remained unchanged in its primary function, and the container to which it is attached – matchbox or tin can – ...18... since the 1850s, which is remarkable when so many extraordinary changes have happened elsewhere ...19....

What were the original functions of the label? ...20..., there was the simple need to say what was ...21.... The shape of the bottle or might already suggest this, but the word 'Strawberry', for example, impressed on the ...22... would identify the ...23... inside it, or the word 'Burgess' ...24.... Probably the most necessary labelling was for medicines, where it was vital that the contents and directions for use ...25... plainly.

The second aim was to glamorise the pack. ...26... or pictorial image, particularly if hand-coloured, would instantly enliven ...27.... Some designs were relevant, depicting a scene in which the product was used; some were purely ...28..., using devices such as a beautiful girl. ...29... colour printing, decorative labels could be mass-produced in a variety of sizes. They could be stuck onto, say, a box of gloves, which instantly transformed it into ...30....

Manufacturers soon noticed that their products ...31... if they had an element of prestige ...32... on the label. The presence of the royal coat of arms, a string of medals won at ...33..., or a testimonial from a respected analyst as to ...34..., gave customers confidence in ...35... what they were buying. By the 1950s, a further sales device was in general use – the direct incentive. Incentives had, indeed, been ...36... since the 1880s when, on Sunlight soap boxes, ...37... was offered to anyone who could find any impurity in the product. Other inducements were offered at that time, including the pack that ...38... after the contents had been consumed, and ...39... that could be saved and stuck into a scrapbook. By the end of ...40..., the promotional pack bearing details of a competition, free gift or price reduction was commonplace.

CAE Reading Comprehension Test

Labels

A The paper label must be one of the most abundant of art forms and, taking into account the number of people it reaches, one of the most influential. Here is a brief look at the history and functions of the labels on everyday products.

Manufacturers have become increasingly aware of how important a part the label plays in the sales of their products. Considerable attention is now given to the visual impact of the label, particularly its overall colour, this being the first point of visual contact. In the early days, brands were sometimes asked for their colour only, to such an extent that the company would incorporate the colour into the brand name. Thus, there was 'Green Label' chutney and 'Red Label' tea.

Manufacturers ought to remember that the label can become so opulent in comparison with the container's contents that the customer's expectations are dashed when the contents are revealed. In 1888, when the cigar market in the US was moving towards ever more lavish labels, an article in the New York Sun commented, 'The label is often better than the cigar...'

B *Since label design plays such an important role in a product's continuity, it is surprising that manufacturers today allow labels to be vandalised by a clutter of competition announcements and price reductions. One can only suppose that the public are not offended by these recent design changes, which are now commonplace in the supermarket – 'the gallery of commercial art' – where some 2,000 designs are displayed. The variety is certainly incredible. There are those designs which follow the latest fashion, those which are modern but not 'over the top', those which retain much of the tradition of previous generations yet are still smart, and those which try to look nostalgic. Design apart, the label itself has remained unchanged in its primary function, and the container to which it is attached – matchbox or tin can – has altered little since the 1850s, which is remarkable when so many extraordinary changes have happened elsewhere in daily life.*

C *What were the original functions of the label? In the first instance, there was the simple need to say what was inside the pack. The shape of the bottle or jar might already suggest this, but the word 'Strawberry', for example, impressed on the side of a jar would identify the type of jam inside it, or the word 'Burgess' the manufacturer. Probably the most necessary labelling was for medicines, where it was vital that the contents and directions for use should be seen plainly.*

The second aim was to glamorise the pack. A design or pictorial image, particularly if hand-coloured, would instantly enliven the overall effect. Some designs were relevant, depicting a scene in which the product was used; some were purely to gain attention, using devices such as a beautiful girl. With the arrival of colour printing, decorative labels could be mass-produced in a variety of sizes. They could be stuck onto, say, a box of gloves, which instantly transformed it into a desirable present.

D *Manufacturers soon noticed that their products sold better if they had an element of prestige attached to them on the label. The presence of the royal coat of arms, a string of medals won at trade exhibitions, or a testimonial from a respected analyst as to a product's*

purity, gave customers confidence in the quality of what they were buying. By the 1950s, a further sales device was in general use – the direct incentive. Incentives had, indeed, been in existence since the 1880s when, on Sunlight soapboxes, a handsome reward was offered to anyone who could find any impurity in the product. Other inducements were offered at that time, including the pack that had a function after the contents had been consumed, and the label that could be saved and stuck into a scrapbook. By the end of the 1950s, the promotional pack bearing details of a competition, free gift or price reduction was commonplace.

Answer the following questions by choosing from the four sections of the extract (A-D). You may choose any section more than once. Please write all your answers (A, B, C, or D) on the *Answer Sheet* in front of the related number.

In which section are the following mentioned?

1. the aspect of a label the people notice before any other
2. labels that have too much information on them
3. a feature of labels that made them more exciting
4. an early example of a practice which later become widespread
5. a reference on a label to awards that a product had received
6. developments in labels that customers do not seem to object to
7. containers that could be used for something else when they were empty
8. people being disappointed in a product after they have bought it
9. a development that made it possible to produce interesting labels in large quantities
10. manufacturers encouraging people to collect labels
11. products that people might buy for other people because of the label
12. a reason why certain products were given the names they were
13. labels that showed the product itself
14. products that were presented as being superior in some way