2.5 2.8 2.2 32 36 2.0 40

28 32 25 36 2.2 2.0

100 mm

1.0
2.8 2.5
32 2.2
36
1.1 40 2.0
1.8
1.25 1.4 1.6

ABCDEFGHIJKLMNOPQRSTUVWXYZ
abcdefghijklmnopqrstuvwxyz1234567890

ABCDEFGHIJKLMNOPQRSTUVWXYZ
abcdefghijklmnopqrstuvwxyz
1234567890

1.0 mm 1.5 mm 2.0 mm

A5

2.0 2.2 40 2.5 36 32 2.8

2.0 40 36 32 2.2 2.8 2.5
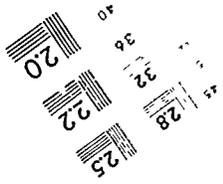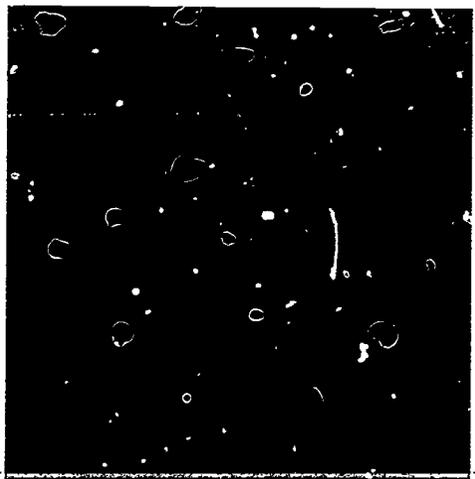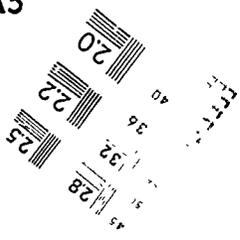
ED 309 644   FL 018 116

AUTHOR   Bensoussan, Marsha; Mauranen, Anna
TITLE   SeDelGap Tests of Macro-Level Reading Comprehension:
                An Exercise in International Collaboration.
PUB DATE   89
NOTE   25p.
PUB TYPE   Reports - Evaluative/Feasibility (142)

EDRS PRICE   MF01/PC01 Plus Postage.
DESCRIPTORS   *Cloze Procedure; Coherence; Cohesion (Written
                Composition); Discourse Analysis; Foreign Countries;
                Higher Education; International Cooperation; Language
                Proficiency; *Language Tests; Multiple Choice Tests;
                *Reading Comprehension; *Reading Tests; *Test
                Construction; Test Reliability; Test Validity

ABSTRACT
                Parallel studies at universities in Israel, Finland,
Canada, and the United States used a "selected deletion gap-filling
test," a variation on the cloze procedure designed to measure reading
comprehension by testing the reader's familiarity with cohesive links
and grasp of text coherence. The test design responded to the growing
demand for a more efficient multiple-choice test of reading
comprehension for large numbers of examinees. The study used
discourse analysis to select the deletions and to examine their
relationships to macro-structures in the text. Pilot testing with the
varied university populations revealed the test to be statistically
satisfactory. Recommendtions are that the test be used as: (1) a
diagnostic test or as a test of language proficiency, with its
deletions adapted to its function; and (2) part of a battery of tests
measuring other aspects of language proficiency. Further research to
expand the scope of the test is under way. (Author/MJE)

SeDelGap Tests of Macro-Level Reading Comprehension

An Exercise in International Collaboration

Marsha Bensoussan

Department of Foreign Languages

Haifa University

Mount Carmel

Haifa, ISRAEL 31999

Anna Mauranen

Language Centre

University of Helsinki

Fabiankatu 26

00100 Helsinki, FINLAND

ABSTRACT

To measure reading comprehension, teachers often ask students to fill in gaps in a text. This basic test format has many variations. Traditionally, the test in which deletions occurred after every nth word was termed 'cloze procedure.' Later, it was called a 'random cloze' test, to be distinguished from the rational cloze, in which the test designer decided which words to omit from the text. To avoid confusion, we have termed our test SeDelGap, 'selected deletion gap-filling test.'

There is controversy about exactly what is measured by a cloze test. Some researchers (Oller, Bormuth, Jonz) have claimed that it is a global measure of reading comprehension, while others (Alderson, Porter, Klein-Braley) have argued that it merely shows a limited knowledge of collocations on the micro-level. A discourse cloze, deleting only cohesive markers (i.e., pronoun anaphora and conjunctions), was described by Levenston, Nir, and Blum-Kulka (1984). The Cohesion Cloze (Bensoussan, forthcoming) is based on the same principle and claims that the blanks are independent of each other.

The SeDelGap principle measures reading comprehension on the macro-level, testing the reader's familiarity with cohesive links and grasp of text coherence. Discourse analysis is used to select deletions and to examine their relations to macro-structures in the text.

To this end, parallel studies at different universities are being carried out with macro-level gap-filling tests. In one series of experiments, tnree texts, designed by Bensoussan and otner researchers (Elizlabeth Tricomi at SUNY Binghamton and Randall Uenara at Harvard Summer Scnool in ESL) were administered to university students in Israel (Haifa University), Finland (University of Helsinki), Canada (OISE), and tne USA (SUNY Binghamton and Harvard Summer Schoolin ESL) during the years 1981-1987. Item analysis showed the tests to be statistically satisfactory.

Test researchers at the University of Helsinki have been experimenting independently with the 'semantic cloze' based on similar principles. Simultaneous testing research in different universities for speakers of many different native languages is exciting. It is noped tnat tne SeDelGap principle can be used to generate more researcn and better tests.

I. INTRODUCTION

Tne SeDelGap Test (Selected Deletion Gap-Filling Test) was designed in response to tne growing demand for a more efficient multiple-choice test of reading comprehension for large numbers of examinees. Most multiple-choice tests are inefficient in that, for tne amount of reading required, tney yield relatively few questions. Moreover, comprenension of tne questions adds an additional component to tne test. Resulting from both text and questions, scores do not necessarily reflect readers' comprehension of the text alone. Questions may also reflect the examiner's interpretation of tne text, thus biasing results. Tne Cloze procedure, although claiming to test reading comprehension, has been criticized for testing readers' micro-level familiarity with a limited range of collocations and idioms rather than examining macro-level comprehension of tne writer's ideas and opinions in tne text. Tne SeDelGap Test aims to combine the multiple-choice and cloze techniques to test reading comprehension on tne macro-level.

A summary of tne development of cloze research would be helpful in explaining tne rationale behind tne SedelGap Test.

II. REVIEW OF THE LITERATURE

A. Cloze Tests as a Measure of Readability

1. Definition of 'Cloze'

Tne term 'cloze' first appeared in an article by Taylor (1953) as a better measure of readability than readability formulas. Taylor recommended random deletion to sample the ability of a reader to comprehend a text.   Bormuth (1966) reports high correlations between readability formulas and cloze passages.

Oller and Conrad (1972) explain the reasoning behind the 'cloze' procedure:

> Tne term 'cloze' was used with the notion of Gestalt "closure" in mind, referring to the natural human psychological tendency to fill in gaps in patterns. The restoration of words deleted from a selection of prose in order for the passage to make sense is a special use of this ability to complete broken patterns. (p. 183)

Carroll (1972) explains Taylor's procedure:

> The procedure involves taking a passage of text and deleting words in it by some rule, e.g., every 5th word, every other noun, or every other "function" word. A subject is then

presented with the passage and asked to guess the missing words. (p. 18)

Although linguistic criteria (parts of speech or function words) may enter into this early cloze procedure, it is the randomness, and not the structure of the text, that counts. From this automatic, mechanical deletion process sprang a whole literature which applied this procedure to a large variety of texts and students, making claims for its performance, criticizing its effectiveness, and suggesting modifications in scoring methods and deletion rates.

2. Reasons for Advocating the Cloze

A number of researchers (Taylor 1953 and 1956, Gilliland 1970, Hirsch 1977) see the cloze procedure as an accurate measure of readability for two reasons: it includes the reader, and it makes use of semantic and syntactic redundancy in the text (i.e., the context) in the calculation of the readability score. That is, it corrects some of the faults of the readability formulas.

Redundancy, as defined by Klare (1963) refers to "the extent to which a given unit of language is determined by nearby units" (p. 172). Like readability, perception of redundancy varies not only with the materials, but also with the readers (Klare 1963, pp. 173-174).

Some researchers claimed that the cloze is a global measure of language proficiency for native speakers of English (Weaver 1962 and

1965, Bormutn 1967 and 1968, Ramanauskas 1972, Oller 1975, and Ozete
1977). Tnis claim was soon extended to include nonnative speakers of
English as well (Oller and Conrad 1972, Oller 1973, Irvine et al.
1974, Stubbs and Tucker 1974, Jonz 1976, Cninara et al. 1977, Berkoff
1979).

Finally, the SeDelGap test yields more items per minute than the
traditional multiple-choice test. Whereas a text of 400 words mignt
yield only ten multiple-choice questions in thirty minutes, it would
yield approximately thirty SeDelGap items in tne same amount of time.
Tnus the SeDelGap test would give the language teacner more
information tnan traditional multiple-choice items about students'
reading comprenension during tne same amount of test time. It would
also increase tne test's reliability since statistical reliability
increases witn tne number of test items.

3. Problems with the Random Cloze Procedure

Not all researchers are entnusiastic about the cloze as a global
measure of reading proficiency, however. Kintsch and Vipond (1977) do
not believe redundancy and readability to be closely related:

The cloze procedure, on the other hand, is probably act.ally misleading. It measures the statistical redundancy of a text, which is a far cry from its comprehensiblity. By that score, a high-order statistical approximation to English that nevertheless constitutes incomprehensible gibberish would be preferred to a well-organized text with less predictable local patterns. (p. 337)

Other researchers are also skeptical of the random cloze procedure (Carroll 1972; Porter 1975; Alderson 1969, 1979, and 1980; Baten 1981; and Klein-Braley 1981).

Opponents of the random cloze present a list of drawbacks. They state that it does not measure what its promoters say it does. Language production being necessary, it is not only a measure of reading ability (Porter 1975). Changes in deletion rates can alter the test unpredictably, so that it cannot be universally applied to every text (Alderson 1969, 1979, and 1980; Klein-Braley 1981). It is not a test of global comprehension across sentence boundaries but a discrete item test that is sentence (or even clause) bound (Alderson 1969, Carroll 1972, Klein-Braley 1981). Random cloze tests do not always distinguish between natives and nonnatives (Alderson 1980) since even natives also have difficulty filling in the cloze and are not necessarily able to get a perfect score (as would normally be expected on a test for foreign language learners).

4. Modified Rational Cloze

Having rejected tne random cloze as not being an automatically valid testing procedure, a number of researcners suggested rational deletion metnods according to linguistic principles (Weaver 1962, Greene 1965, Alderson 1969, Cranney 1972-73, Klein-Braley 1981, Bensoussan and Ramraz 1984).

Tne rational cloze procedure can be used as a measure of text difficulty for a certain population. Depending on the placement of blanks, different kinds of tests can be obtained. Blanks can be put in place of content words, function words, parts of speech, markers of conesion; these words can be used to test comprehension on the micro-level or the macro-level. Even tne rational cloze does not necessarily test the student's grasp of tne content or ideas in the text, nowever.

Greene (1965) explains the rationale behind a modified cloze test wnicn he constructed:

each possible deletion was evaluated by the author for possible effectiveness and deletions made on this rational rather than mechanical basis. For each word deleted under the modified cloze procedure, there was felt to be sufficient redundancy remaining in the passage so that a superior reader could make positive identification of the missing word. (pp. 213-214)

Other researchers advocate deleting certain parts of speech (Weaver 1962, Klein-Braley 1981) or a certain percentage of content vs. function words (Berkoff 1979).

Working with nonnatives, Bachman (1982) deleted on the basis of syntactic (clause-level context), cohesive (inter-clause or inter-sentential context), or strategic (parallel) patterns of coherence (p. 63). Also working with EFL students, Berkoff (1979) and Sim (1979) experimented with rational cloze to test comprehension of items of coherence and cohesion.

Using the rational cloze to measure reading comprehension, the researcher could use Greene's (1965) criteria in determining deletions. The resulting cloze tests should contain sufficient redundancy to make sense to the competent native reader.

A multiple-choice modification of the rational cloze procedure was designed by Bensoussan and Ramraz (1984). The basic advantage of this method over the multiple-choice test is that the correct answer does

not reflect tne tester's interpretation of tne text, but is an integral part of the text. Tne basic advantage over standard cloze procedure is that the focus is on recognitioa, not production. That is, tne focus is on reading uncontaminated by tne element of writing.

5. Cloze Procedures on the Discourse Level

Recently there have been two efforts at tapping comprehension on tne discourse level by means of a modified cloze procedure. Tne Discourse Cloze (Levenston, Nir, and Blv'n-Kulka 1984) deleted only overt cohesion markers of co-reference and connectives between propositions. Tney assumed that the correct completion of macro-level items deleted from a text indicates understanding of the whole discourse.

Tne second effort is the Semantic Cloze (Mauranen 1988) to test comprehension of advanced, academic texts. It aimed to delete more macro-level than micro-level words, deleting content words (nouns, verbs, adjectives, and adverbs) and discourse markers, also including alternative responses for each deletion. It has the advantage over a standard cloze of actually avoiding the micro-level bias. There is also evidence (Mauranen 1988) to suggest that the test is sensitive to cnanges in students' reading comprenension skills.

III. SeDelGap TEST PRINCIPLES

The SeDelGap test measures reading comprehension on the macro-level, testing both the reader's familiarity with cohesive links and grasp of text conerence. Discourse analysis is used to select deletions and to examine their relations to macro-structures in the text.

A. Selection of Texts

The cnoice of text nas an important effect on gap selection and scores. In courses of English for Academic Purposes (EAP), general expository texts are used wnicn test students' ability to follow a logical argument (i.e., recognition of general ideas versus detail, cause versus effect, agreement versus disagreement/reservations). In courses of specialized uses of English such as business or chemistry (English for Special Purposes: ESP), scientific texts would require students, in addition to following an argument, to be acquainted with specialized vocabulary, specific procedures, and sequencing of sentences (e.g., hypothesis/proof, definition/givens/procedure/conclusion). Literary texts, moveover, wnicn are more freely organized demand different reading skills from the reader (e.g., flasnbacks, descriptions, comments, and metaphorical language).

Expository texts used in the SeDelGap procedure should be culturally and tnematically neutral but still interesting. They must be culturally neutral so as not to disadvantage any particular group of students. Tnematic neutrality ensures objectivity; when students read a text that clashes with tneir own personal, religious, or political opinions, they become too impatient or irritated to fully comprenend tne writer's point of view. Instead, they tend to impose their own opinions or schemata on the text. If the text is dull, on the other hand, students become bored. All these deviations interfere witn an objective assessment of students' reading comprehension and generally result in artificially low test scores.

The texts should also present new information to the students, even if it is in a familiar content area. Tnis is not only to avoid boredom, but to tap the students' ability tc cope with new meanings expressed in a foreign code.

B. Selection of Deletions

A blank may take the place of one or more words so that the meaning
is recoverable from the text. Textual clues derive from redundancy,
collocations, denotations and connotations (negative and positive),
and opposites in tne text. Each blank in the text should be
conceptually and linguistically independent of the others.

To restore a deletion, students should have to read beyond the
clause in which the deletion appears. The SeDelGap procedure is based
on deletions which can be completed by knowledge of the ideas and
structure of tne text on the macro-level. It tests familiarity with
the coherence and cohesion of the text. Logical relations of
coherence would include general/specific (e.g., example),
cause/effect, contrast/comparison, addition, series, parallel ideas,
analogy/metaphor/simile. Tne writer's attitude or intention is
anotner macro-level construct that should be tested. Cohesive markers
include pronouns (e.g., it, hers, this), substitution (one, do, so,
not), sentence connectors (conjuctions), and lexical cohesion:
repetition, near-synoryms, superordinates and subordinates. Deletions
wnich can be completed by micro-level knowledge of the immediate words
surrounding the text are more likely to be testing grammar than
comprehension of ideas, and are best avoided.

Wnen the SeDelGap procedure also includes the element of
multiple-choice distractors, each deletion must be able to yield
alternate answers. Tne blank may be situated in such a way as to test

15

an idea, pattern, or structure directly (i.e., the meaning of a particular word) or indirectly (i.e., a word tapping a logical relation such as contrast, exemplification, cause/effect).

C. Selection of Multiple-Choice Alternate Responses

Each deletion should test only one point of comprenension (e.g., idea, verb tense, sequencing, contrast, content/function). It is desirable for all responses to be parallel in form (i.e., all adjectives, gerunds, conjuctions, etc.) and register. Opposites can be good distractors.

All alternate responses must be grammatically correct since the SeDelGap does not test grammar but reading comprehension. To find the correct response, the reader needs to make use of the context surrounding the blank.

Since this is a foreign language test, it is best to avoid unusual words, and fine semantic and syntactic distinctions which may confuse even the native speaker (e.g., its vs. it's).

IV. COMPARISON OF SeDelGap TESTS

Recently, parallel studies at different universities have been carried out with macro-level gap-filling tests. In one series of experiments, three texts, designed by Bensoussan and other researchers (Elizabeth Tricomi at SUNY Binghamton and Randall Uehara at Harvard Summer School in ESL) were administered to university students in Israel (Haifa University), Finland (University of Helsinki), Canada (OISE), and the USA (SUNY Binghamton and Harvard Summer School in ESL) during the years 1981-1987. Item analysis showed the tests to be statistically satisfactory (see Table).

TABLE: SeDelGap Test Collaboration

Text	Place	Year	No. Stud	No. Quest	Reliab	Mean Score	Standard Error
Balloons	Harvard	(85)	60	40	.88 (KR)	65 %	2.5
Balloons	Harvard	(85)	74	40	.89 (KR)	69 %	2.4
Rich vs. Poor	Haifa	(81)	73	21	.80 (KR)	63 %	---
Rich vs. Poor	OISE	(87)	59	21	.65 (Hoyt)	62 %	1.9
Rich vs. Poor	OISE	(87)	54	20	.73 (Hoyt)	55 %	1.9
Rich vs. Poor	OISE	(87)	52	20	.66 (Hoyt)	55 %	2.0
Grant and Lee	SUNY	(85)	82	34	.76 (KR)	68 %	2.4
Grant and Lee	Harvard	(85)	104	34	.88 (KR)	66 %	2.3
Grant and Lee	SUNY	(86)	95	34	.88 (KR)	67 %	2.3
Grant and Lee	Helsinki	(86)	97	34	.83 (Alfa)	70 %	---
Grant and Lee	SUNY	(86)	83	30	.84 (KR)	64 %	2.3

18

V. FURTHER IMPLICATIONS FOR TEACHING AND TESTING LANGUAGE

Tne SeDelGap may be used as a diagnostic test or as a test of language proficiency. Tne test designer can select blanks and alternate responses according to the function of tne test.

Wnat tne SeDelGap, in this form, does not test is readers' independent and critical interpretation of tne main points of a text. To remedy this situation, we nave been experimenting with a combination of question types, with ordinary general comprehension questions at the beginning and/or end of tne undeleted part of the text, and deletions in most of the middle part of the text. This mixed format nas worked well and has not been confusing to students.

Tne SeDelGap would be useful in a battery of tests, also including testing formats sucn as written summary and oral interview, each focusing on a different aspect of language proficiency.

Simultaneous testing research in different universities for speakers of many different native languages is exciting. It is noped tnat the SeDelGap principle can be used to generate more research and better tests.

REFERENCES

Alderson, J. Charles (1969) "Tne Cloze Procedure and Proficiency in Englisn as a Foreign Language," TESOL Quarterly, 13 (2), 219-227.

---- (1979) "Tne Effect on the Cloze Test of Cnanges in Deletion Frequency," Journal of Research in Reading, 2 (2), 108-119.

---- (1980) "Native and Nonnative Speaker Performance on Cloze Tests," Language Learning, 30 (1), 59-76.

Bachman, Lyle F. (1982) "Tne Trait Structure of Cloze Test Scores," TESOL Quarterly, 16 (1), 61-70.

Baten, Lut (1981) Text Comprehension: Tne Parameters of Difficulty in Narrative and Expository Prose Texts: A Redefinition of Readability. Unpublished Pn.D. Tnesis. University of Illinois, Urbana, Illinois.

Bensoussan, Marsha and Rachel Ramraz (1984) "Testing EFL Reading Comprehension Using a Multiple-Choice Rational Cloze," Modern Language Journal, 68 (3), 230-239.

----- (forthcoming) "Redundancy and tne Cohesion Cloze" The Journal of Research in Reading.

Berkoff, Nelson (1979) "Reading Skills in Extended Discourse in English as a Foreign Language," Journal of Research in Reading, 2 (2), 95-107.

Bormuth, John (1966) "Readability: A new approach," Reading Research Quarterly, 1 (3), 79-132.

---- (1967) "Design of Readability Research," in J. Allen Figurel (ed.) Proceedings of the Eleventh Annual Convention, International Reading Association, Vistas in Reading (Vol. II, Part I). Eastern Michigan University: International Reading Assoc., Inc., 485-489.

---- (1967) "Comparable Cloze and Multiple-Choice Comprehension Test Scores," Journal of Reading, 10, 291-299.

---- (1968) "Cloze Test Readability: Criterion Reference Scores," Journal of Educational Measurement, 5 (3), 189-196.

---- (1970) On the Theory of Achievement Test Items. Chicago and London: University of Chicago Press.

Carroll, John B. (1972) "Defining Language Comprehension: Some Speculations," in Freedle and Carroll (eds.) Language Comprehension and the Acquisition of Knowledge, 1-29.

Cninara, Tetsuro, Jonn Oller, Kelley Weaver, and Mary Anne Cnavez-Oller (1977) "Are Cloze Items Sensitive to Constraints Across Sentences?" Language Learning, 27 (1), 63-73.

Cranney, A. Carr (1972-73) "Tne Construction of Two Types of Cloze Reading Tests for College Stuents," Journal of Reading Benavior, 5 (1), 60-64.

Gilliland, Jack (1972) "Tne assessment of Readability -- an Overview," in Melnik and Merritt (eds.) Tne Reading Curriculum, 144-158.

Greene, Frank P. (1965) "Modification of tne Cloze Procedure and Cnanges in Reading Test Performances," Journal of Educational Measurement, 2 (2), 213-217.

Hirsch, Eric Donald (1977) The Philosophy of Composition. Chicago: Univ. of Cnicago Press.

Irvine, Patricia, Parvin Atai, and John W. Oller (1974) "Cloze, Dictation, and the Test of English as a Foreign Language," Language Learning, 24 (2), 245-253.

Jonz, Jon (1976) "Improving on the Basic Egg: tne M-C Cloze," Language Learning, 26 (2), 255-265.

Kintsch, Walter and Douglas Vipond (1977) "Reading Comprehension and Readability in Educational Practice and Psychological Tneory," in Nilsson (ed.) Perspectives, 329-365.

Klare, George (1963) The Measurement of Readability. Ames, Iowa: Iowa State University Press.

Klein-Braley, Cnristine (1981) Empirical Investigations of Cloze Tests: An Examination of the Validity of Cloze Tests as Tests of General Language Proficiency in English for German University Students. Unpublished Pn.D. Thesis. Duisburg University, Duisburg, Germany.

Levenston, Edward, Raphael Nir, and Snoshana Blum-Kulka (1984) "Discourse Analysis and tne Testing of Reading Comprehension by Cloze Techniques," in A. K. Pugh and J. M. Ulijn (eds.) Reading for Professional Purposes, 202-212.

Mauranen, Anna (1988) "Tne Semantic Cloze: An Approach to Testing EFL Reading." Paper presented at TESOL conference, Cicago, Ill.

Oller, John W. (1973) "Cloze Tests of Second Language Proficiency and Wnat They Measure," Language Learning, 23 (1), 105-118.

—— (1975) "Cloze, Discourse, and Approximations to English," in Burt and Dulay (eds.) TESOL 1975, 345-55.

—— and Christine A. Conrad (1972) "The Cloze Technique and ESL Proficiency," Language Learning 21 (2), 183-195.

Ozete, Oscar (1977) "The Cloze Procedure: A Modification," Foreign Language Annals, 10 (5), 565-568.

Porter, Don (1975) "Modified Cloze Procedure: A More Valid Reading Comprehension Test," ELT Journal, 30 (1), 151-155.

Ramanauskas, Sigita (1972) "The Responsiveness of Cloze Readability Measures to Linguistic Variables Operating over Segments of Text Longer than a Sentence," Reading Research Quarterly, 8 (1), 72-91.

Sim, Donald (1979) Links Between Context and Selected Features of Grammatical Cohesion in English, and Performance in Advanced Reading Comprehension for Overseas Students . Unpblished Pn.D. Tnesis. University of Manchester.

Stubbs, Joseph and G.R. Tucker (1974) "The Cloze Test as a Measure of English Proficiency," Modern Language Journal, 58 (5-6), 239-242.

Taylor, Wilson L. (1953) "Cloze Procedure, a new tool for measuring readability," Journalism Quarterly, 30, 415-433.

----   (1956) "Recent Devlopments in the Use of the Cloze Procedure," Journalism Quarterly, 33, 42-48,99.

Weaver, Wendell (1962=1977) "Tne Predictability of Omissions in Reading and Listening," in Albert J. Kingston (ed.) (1977) Toward a Psychology of Reading and Language: Selected Writings of Wendell W. Weaver. Athens, Georgia: University of Georgia Press. Pp. 4-8.