Presidential Address

THE CLOZE PROCEDURE REVISITED

EARL F. RANKIN
University of Kentucky

Sixteen years ago I had the pleasure of reading a paper entitled "The Cloze Procedure — Its Validity and Utility" at the 1958 meeting of the National Reading Conference in Fort Worth Texas. That occasion, so far as I am aware, was the first time that any research or theory on the cloze was presented at a national meeting of people in the field of reading. Since that time, work utilizing the cloze technique in both basic and applied research has mushroomed dramatically. It is my purpose today to look at the cloze in the perspective of my initial research in the late 1950's and to consider several subsequent developments in cloze research. I would also like to mention some limitations of the cloze procedure which may not be generally recognized. Before continuing, however, let me pause to reminisce about the development of my dissertation which formed the basis for my paper at the NRC meeting.

At the University of Michigan in 1956, I became interested in "information theory" and its possible applications to the study of reading. It seemed to me that the concept of "redundancy" in written language should have some relationship to readability. In my research through the literature on information theory, I soon became bogged down in the mathematical complications of research on this topic. One day in the library, while looking through the *Psychological Abstracts,* I came to a reference on something called the "cloze procedure" by Wilson Taylor. It was immediately evident to me that Taylor had done essentially what I had been trying to do, only he had accomplished this some three years earlier. As I look back upon this time, the main thing that I remember is a great sense of relief, because I saw the possibility of using the cloze procedure in my dissertation without having to include a review of research on information theory in the dissertation. I realized the potentialities of cloze for research and reading and believed that no one in the field of reading had published anything on this topic. It turned out that Marion Jenkinson completed her dissertation on the cloze technique at the University of Chicago the same year that I completed mine. These two studies, I believe, were the first cloze studies that explored, in some depth, several aspects of reading comprehension. At that time, Taylor's publications were concerned chiefly with the use of cloze as a measure of readability of prose passages.

After I had completed my dissertation, Don Smith, the chairman of my doctoral committee at the University of Michigan, told Oscar Causey, the president and founder of the National Reading Conference, about my work, which led to my being invited to deliver an address at the next NRC meeting. This happenstance, in turn, resulted in my being offered a professorship at TCU. I arrived in Fort Worth shortly after Oscar's death to become acting head of the NRC, thus beginning a long association with this organization which led, indirectly, to membership on the Board of Directors and, later, nomination for the vice-presidency. But enough of this nostalgia. Let me return to my theme.

Keeping in mind that Taylor developed the cloze technique in 1953, it is interesting to note the development of research studies using this technique since that time. In 1958, when I presented my paper in Fort Worth, there were about eight references to publications using the cloze. In 1964, when I presented a survey of research on cloze at the National Reading Conference meeting in Dallas, this number had grown to 58. Charles and Rachel Bickley of Francis Marion College and I have accumulated a listing of 252 research papers using the cloze procedure from 1953 to the present. If we assume, for the moment, that 252 is a fairly accurate count, then it would follow that only about 23% of the total number of studies had been carried out by 1964. That would leave approximately 77% of the total number of cloze studies to be published during the past nine years and 48% published since 1968. A glance at a frequency distribution based upon cloze studies which have been published between 1953 and the present dramatically reveals an extensive use of the cloze procedure as a research tool.

Since the 1958 meeting, about 40 doctoral dissertations have been carried out using the cloze procdure. Looking at articles published in the NRC yearbooks since that time, I find about 25 publications using the cloze. A large number of NRC members have worked with the cloze procedure. Names such as Wendell Weaver, Al Kingston, Larry Hafner, Frank Greene, Jaap Tuinman, and Dick Bloomer come to mind. Many others, of course, could be mentioned. It has been my pleasure to have delivered five research reports using cloze at our meetings since 1958.

Areas of investigation, as revealed in cloze publications since 1958, include readability, reading comprehension, learning, information, redundancy, thinking, numerous language variables, teaching, aptitude, readiness, listening, flexibility, and context clues. There are obviously many sub-categories within these areas which could be mentioned. This list reveals the broad spectrum of reading-related topics which have been explored through the use of the cloze technique.

Comprehensive reviews of the cloze procedure have appeared in Rankin's paper in the *8th NRC Yearbook* (1959), Rankin's Survey in

the *14th NRC Yearbook* (1965), Potter's review in a Southwest Regional Laboratry paper (1968), the conspectus by Bickley and associates in the *Journal of Reading Behavior* (1970), and Jongsma's reviews in the *Occasional Papers in Reading* series at Indiana University (1970) and the *ERIC-IRA Reading Information Series* (1971).

Let me return to a consideration of several findings and suggestions for future use of the cloze which are to be found in my dissertation or address to the National Reading Conference in 1958. The major findings in these studies were as follows: (1) The cloze procedure can be used to measure (a) specific reading comprehension for a particular passage or (b) general reading comprehension skills as measured by standardized reading tests. (2) It can be used in the measurement of comprehension to place relative emphasis upon comprehending "lexical" or "structural" meaning. (3) It can be used to measure both pre- and postreading knowledge. (4) It can be used to measure the increase in knowledge resulting from reading by computing "gains" between pre- and postreading cloze tests. (5) An objective scoring system results in measurements which are as valid as a "subjective" (i.e., synonym) scoring system. (6) Prereading cloze tests are more susceptible to the influence of some personality effects than postreading cloze tests. At that time, I made several sugestions for future study and use of the cloze procedure. These suggestions included: (1) the study of gain scores as a measure of "learning through reading," (2) the use of postreading cloze tests to reduce the effects of personality variables, (3) the study of different scoring techniques, (4) the study of cloze results obtained at different age-grade levels, (5) a caution about anxiety reactions to initial item difficulty and a suggestion for gradually eliminating the amount of context in the construction of a cloze test or exercise from the beginning to the end of the passage, (6) the use of diagnostic patterns based upon prereading cloze scores, postreading cloze scores, and cloze gain scores for clinical studies, (7) the use of cloze as a teaching technique, (8) the use of cloze as a measure of readability for specific passages read by particular groups of readers.

Looking back upon cloze research related to these points in my work in the late 1950's, I would like to make a few comments upon subsequent trends. Despite the finding that different aspects of comprehension could be measured by restricting the deletion of words to different form classes, most of the work using cloze has made use of the "every nth" deletion which measures "general" comprehension skill more than it measures "specific" comprehension of a particular passage. In my opinion, the almost exclusive reliance upon this type of cloze has strengthened the influence of general verbal abilities and intelligence upon the cloze measurement of reading comprehension. The fact that

the "every nth" cloze correlates better with the results of standardized reading tests probably accounts for the greater use of this measurement procedure in the literature. A comprehension measurement which decreases the influence of these extraneous variables and increases the influence of information contained in the passage read upon the measurement process should be studied more carefully.

My distinction between two methods of constructing cloze tests to measure "lexical" versus "structural" comprehension of meaning has been widely misunderstood. I hypothesized that a restriction of deletions to form classes such as nouns and verbs would result in a measurement which would place greater emphasis upon the understanding of the substantive content of a passage than it would upon the syntax of a passage. I also predicted that the total amount of structural meaning in a passage would be reduced more than the total amount of lexical meaning, if cloze tests were constructed by deleting every "nth word." Hence, such a test should be a better measure of structural than lexical comprehension. This hypothesis was based upon the observation that several *different* clues such as function words, morphological clues, and word order signal structural meaning in our language. Since a random or "every nth" word relation should sample *all* of these clues and still leave many nouns, verbs, adjectives, and adverbs (which occur in great abundance) in the remaining context to signal lexical meaning, an "every nth" deletion should reduce the total amount of structural meaning more than the total amount of lexical meaning. At no time did I maintain that either type of cloze was a pure measure of either aspect of meaning. My results confirmed these hypotheses. Since that time, the practice of deleting "function words" only has been widely used to measure structural comprehension and such usage has been attributed to my early work. To my knowledge, there is no proof that the use of a function word deletion system results in a better measurement of structural meaning than an "every nth" deletion. Many other clues to structure would not be sampled by a function word deletion system. No one has ascertained the relative extent to which different clues to structure serve to signal structural meaning in our language. This problem needs to be studied carefully in order to improve our measurement and understanding of these important linguistic variables. In the meantime, it seems that a cloze deletion system which samples the small number of language elements which occur with very high frequency to signal structural meaning, and also produces only a small sample of the larger number of language elements which occur quite infrequently to signal lexical meaning, has much to be said for it as a measure of structural meaning. If we expand the concept of structure to extend beyond sentence boundaries, then it might follow that the use of a sampling system not restricted to sentence constraints might measure

a grasp of structure in a broader context than the sampling of function words, alone, would permit.

Most of the work in my initial cloze research was based upon lexical deletions. These deletions produced good measures of both prereading and postreading knowledge and reading gains. The study of gains appeared to me at the time to offer tremendous potentialities for measuring reading as a learning process rather than an end product. I obtained highly significant differences between pre- and postreading cloze tests both based upon the same group of subjects, before and after reading, and two comparable groups of subjects, one receiving a prereading test and the other receiving a postreading test. Today, the most often quoted finding concerning cloze gains is based upon the work of Coleman and Miller, who failed to obtain significant gains using cloze. Both the work of Taylor and myself, which reported highly significant gains, has been largely ignored. It may well be that the use of an "every nth" deletion by Coleman and Miller rather than a lexical deletion might be responsible for their failure to find significant gains. If we consider the fact that the lexical deletions may result in more reading related items than "every nth" deletions and the fact that "every nth" deletions are more subject to the influence of intelligence, general reading comprehension and general language skills than lexical deletions, then it would follow that the use of this deletion system is less likely to result in significant gains than the use of the measurement technique which emphasizes the measurement of lexical comprehension. More work along these lines might prove productive.

An incidental finding in my early work was that a personality variable, introversion/extraversion, had a significant effect upon prereading cloze tests but not upon postreading cloze tests. Also, I found evidence suggesting that readers with a high level of anxiety tended to do more poorly on a particular precloze test form than on other precloze test forms. This difference may have been due to the greater difficulty on initial items on that form. Both of these personality effects appeared on prereading cloze tests. Subsequent work that I carried out, following up the implications of the first finding, showed that introverts performed differently from extraverts on other standardized reading tests as well as precloze tests. An important implication of this work is that postreading cloze tests, unlike both precloze tests or other reading tests, do not seem to be influenced by some personality variables. Virtually all research has been done on precloze tests over the years. I suggest that the postcloze test needs investigation, despite the somewhat greater length of time required to administer such a test.

Perhaps due to the difficulties entailed in measuring gains, my suggestion concerning the use of a diagnostic pattern based upon cloze tests to measure a combination of prereading knowledge, postreading

knowledge, and gains between pre- and postreading has not been followed up. Also, my idea about constructing a cloze test to eliminate anxiety reactions by gradually fading the amount of context from the first to the last of a test, has not been tried. Despite the difficulties involved, the measurement of improvement as distinct from pre- and postreading knowledge is too important not to be studied. Also, anything that we can do to reduce contamination of measurements by personality variables deserves careful study.

I am glad to say that work concerning the use of cloze over a broad spectrum of age-grade levels has been done. It can be concluded that cloze exercises or tests can be administered to readers from grade one through the adult years. It should be noted that cloze materials for first graders have been modified to make it possible for them to cope with this type of task.

Concerning my suggestion for the use of cloze as a teaching technique, it is very disappointing to see the lack of positive findings concerning the effectiveness of cloze in a teaching context. As Jongsma has observed, most of the research on cloze teaching has been quite deficient in that the cloze has often been used in a mechanical manner which ignores the crucial elements necessary for effective teaching. In any case, we can conclude that giving cloze exercises will probably not result in any improvement as measured by standardized reading tests. Perhaps if we used some more specific measure of a skill, like the use of context clues, we might obtain significant results. In any case, if we keep in mind the admonition that materials don't teach, but rather, provide practice for things that have been taught, then there still may be some hope for the truly effective use of cloze as part of a total instructional setting.

Since the cloze procedure was originally conceived by Taylor as a measure of readability, it is not surprising that work in this area has proved to be more successful than most other areas. Although there may be problems of reliability, etc., when measuring results for individuals, the measurement of readibility is based upon the mean of cloze scores over groups of individuals and, hence, is not subject to these limitations. A basic difference between cloze readability and readability as measured by various formulas has not always been made clear. As Professor Clare and his colleagues at Ohio University have observed, whereas readability formulas "predict" the difficulty of materials for intended readers, cloze tests actually "measure" the difficulty of materials for these readers. Measurements are, of course, far more accurate than predictions. The excellent work of Bormuth at the University of Chicago has demonstrated the value of cloze as a measure of both independent and dependent variables in readability research.

Another area of investigation in which the cloze procedure has

been most fruitful is the study of language. Here, the cloze procedure combined with the insights of the growing science of linguistics has proved to be a basic research tool of great importance. However, this important area of investigation was only touched upon in my initial work which was done about the time that linguistics was only beginning to emerge as an important science in the study of reading.

Before closing, let me mention a few limitations of the cloze which have not always been made clear in the literature. Unfortunately, when some new development comes along, all of us are tempted to emphasize its strengths and to de-emphasive its weaknesses. Let us try to take a more objective look at the cloze procedure in light of both its assets and limitations.

One, no completely satisfactory method has been developed to measure the lexical and structural meaning without producing a confounded measurement or an incomplete measure of either aspect of meaning. Two, in the original concept of cloze by Taylor, cloze was differentiated from the traditional incomplete sentence by its use of some form of mechanical deletion. I wonder if the use of rational deletions or commonality scores (as used by McLeod and others) is in keeping with the distinctive concept of cloze. Three, the use of cloze materials without feedback and other important aspects of teaching does not improve reading comprehension, at least as measured by conventional tests. Four, cloze tests contain many items for which there are no context clues. These items will not often be answered on a precloze test. To the extent that they are not answered, they are not discriminating items from the standpoint of classical test theory. Five, many items in cloze tests are not reading related items. They may reflect background information or general language ability. Six, while cloze tests have empirical validity, they are lacking in face validity. This may be a problem in using cloze too frequently. Seven, although objective scoring may have no effect upon cloze test validity, it may have an effect upon the student's motivation if he learns how the test is scored. Eight, although every-fifth-word deletions are commonly used for most research, there are some indications that this deletion is not suitable for everyone. Nine, young children (i.e. first graders) should not be given cloze materials without some modification of the technique. Ten, although precloze materials are most often used, they have some limitations which postcloze tests do not seem to have. Eleven, cloze tests measure comprehension in a global sense and do not yield the kinds of diagnostic information that many other reading tests provide. Twelve, cloze responses may depend too much on short-range constraints. Presumably, other kinds of constraints operate in ordinary reading comprehension. Thirteen, the conventional fifty-item cloze test does not always result in high enough reliability for individual use.

Longer tests should sometimes be used. Fourteen, conventional cloze tests tend to measure "convergent thinking." To measure "divergent thinking" some modification needs to be made, such as that devised by Byrne, Feldhusen and Kane. Fifteen, cloze tests are better measures of the readability of passages than of reading comprehension. Sixteen, performance on a cloze test, as on any other test, is influenced by the reading ability of the reader and the difficulty of the materials. In addition, performance on the cloze test is influenced by the type and number of items deleted. Until we know more about the possibile interrelationships of these variables and their impact upon cloze measurement, we should be cautious in interpreting cloze tests, particularly when there is a large difference between the reading level of the student and the reading difficulty of the materials upon which the test is based. Seventeen, raw scores on cloze tests have all the disadvantages for interpretation as do raw scores on any other test. Only a few studies have been done in establishing criterion-referenced scores for interpreting cloze results. All of these studies have been based upon every-fifth-word deletions, objectively scored. The use of other deletion ratios or other scoring procedures limits the use of obtained criterion scores in interpreting all cloze test results.

Despite all of the previously-mentioned limitations, the cloze procedure has much to recommend it as a basic technique for studying language and reading. Unlike other measurement procedures, it eliminates the influence of test items upon validity. Furthermore, cloze tests are very easy to construct and to score. The increasingly widespread use of cloze measurement in research studies on many topics suggests its basic utility. Wendell Weaver used the cloze in most of his many research studies throughout his professional career. For me, this alone is ample testimony to the value of the cloze as an indispensable tool for the serious scholar in language and reading. Let us hope that it will always be used carefully with due regard to its inherent limitations as well as its assets.