10-1-1999

Applications of cloze procedure to reading assessment in special circumstances of literacy development

Norbert Francis
Northern Arizona University

Follow this and additional works at: https://scholarworks.wmich.edu/reading_horizons

Part of the Education Commons

Recommended Citation
Applications of cloze procedure to reading assessment in special circumstances of literacy development

Norbert Francis
Northern Arizona University

ABSTRACT

This study forms part of an on-going investigation of second language reading development of elementary school children from a bilingual, indigenous, community where high levels of proficiency in both languages are common among all age groups. Modifications were applied to the standard rational deletion cloze procedure to adapt assessment to the special sociolinguistic circumstances of reading in an indigenous language, and for the purpose of exploring practical classroom applications of the cloze procedure for instruction.

INTRODUCTION

With the growing interest in literacy among speakers of minority languages in countries where the national language holds a virtual monopoly over writing, teachers and researchers working in the area of second language (L2) reading have begun to experiment with locally developed assessment approaches. Exploring the possibilities, for example, of teaching an indigenous language in school represents a major departure from traditional literacy teaching policies; proceeding to design assessment tools to actually evaluate student performance in the formerly excluded language reflects a superior level of consciousness and commitment to the goals of additive bilingualism and biliteracy development. Researchers in Latin America have pointed out that biliteracy, the use of indigenous languages in reading and writing in school, in addition to Spanish, is a necessary condition for the development of both the vernacular and the national languages, particularly in regard to reversing the
erosion and loss of the former (Cerron-Palomino, 1993; Jung, 1992; Chiodi, 1992). Thus, biliteracy and additive bilingual development are necessarily complementary. In all such cases, however, educators face the special circumstances of both bilingual development itself (in relation to the significant social discontinuities and imbalances that characterize the contact between the languages spoken in the community), and reading and writing in the vernacular.

The purpose of this report is to evaluate the suitability of one particular assessment technique, the cloze procedure, for literacy teaching in an indigenous language. On the one hand, classroom-based research of reading development in vernacular languages promises to shed new light on the broader theoretical problems of assessment in bilingual and multicultural contexts. In addition, informal, teacher-designed assessment instruments (cloze being one relatively accessible option of the paper and pencil variety) offer immediate solutions to the problem of the availability of classroom literacy materials in vernacular languages.

THE SOCIOLINGUISTIC CONTEXT OF LITERACY ASSESSMENT

The present study will report on initial methodological considerations that emerged from one particular application of the cloze procedure as it was piloted with a group of 3rd and 5th grade bilingual students. The investigation was designed as a follow-up probe of the limits, so to speak, of cloze testing after its successful application in a previous study in the same school under significantly more controlled, and interactive conditions (Francis, 1994; Francis and Nieto, 1996). Participants in the study were 3rd and 5th grade students from an indigenous community in Central Mexico distinguished among neighboring towns for the high levels of conversational fluency that school-age children exhibit in both Spanish and Na’huatl. Virtually all children beyond the second grade demonstrate, at least, basic, interpersonal, oral comprehension skills in both languages.

However, while a large percentage of entering kindergarten and first graders are monolingual or dominant in Na’huatl, all literacy instruction is in Spanish. Thus, for many students, the regular reading program represents a L2 literacy acquisition situation. For teachers experimenting with vernacular literacy, reading instruction would involve a biliterate enrichment experience for the majority of the bilingual students, and an analogous L2 literacy program for Spanish dominant children. For the latter group, the sociolinguistic context of L2 literacy is clearly of a qualitatively different nature. For monolingual Na’huatl speakers,
Applications of cloze literacy instruction in the vernacular would offer the obvious benefits of L1 literacy development.

As the results reported below will suggest, cloze testing appears to represent a viable alternative (among others) for situations where there are neither generalized or stable literacy practices in the vernacular, and where assessment of literacy in the language has never been attempted. Before outlining the procedures and discussing problems of method, a brief review of the sociolinguistic circumstances of this particular testing program are in order. As is characteristic of many vernacular or indigenous speech communities: 1) except for isolated pilot projects, Na'huatl plays virtually no direct role in the primary grades in literacy instruction, regardless of the range of competence levels that students exhibit in Spanish; 2) for many, if not the great majority of students, the activity itself, of reading a continuous text in the indigenous language represents a completely novel experience; 3) points 1 and 2 reflect a general distribution of language use functions (and for many bilingual speakers a normative allocation) where texts in Na'huatl are rare and hard to find. (e.g.: bibles or religious tracts in private possession, unused first grade bilingual education materials stored away at the local elementary school); 4) the progressive erosion of the vernacular continues unabated in the face of the expansion of the national language into previously traditional/indigenous domains; 5) related to all of the above, a lack of orthographic standardization and a perceived dialectical fragmentation by Na'huatl speakers themselves, call into question, for many, the very concept of vernacular literacy.

Historians of literacy in Mexico, however are familiar with the linguistic and language policy background to the present day situation that should be cause for reflection for researchers of reading and writing development in bilingual contexts. Na'huatl, for example, is one of the indigenous languages of the Americas that was the medium of a great literary tradition that spanned hundreds of years, including both the pre-Conquest epoch and, in alphabetic form, the first period of Spanish colonial rule. Particularly during the 16th century, Na'huatl played a central role in the domains of higher education (at the University of Mexico and other institutions sponsored by the church), literacy teaching and writing in general (including not only religious texts, but legal documents, ethnographies, and historical accounts), and the theater (Heath, 1972, Lockhart, 1992, Sten, 1982, Leon-Portilla, 1996). Given the precipitous ascendancy of Na'huatl as the lingua franca of evangelization and public administration during the period following the fall of the Aztec empire, it
is entirely possible that in many regions of Mexico more people were literate in Na'huatl than in Spanish.

Thus, one of the great ironies of the history of literacy is that today, among present-day speakers of the language (numbering between one and 1.5 million) that learning to read and write is generally associated with learning Spanish as a second language, and that the voluminous body of literature in their language, catalogued in libraries and universities, is primarily the object of study by historians, linguistics, and anthropologists. Nevertheless, in the speech communities themselves, the historical consciousness of this literary tradition and the knowledge of the continued existence of an accessible archive that embodies it, continues, in turn, to spark interest in the development of the Na'huatl language, as well as many of the other 56 autochthonous languages of Mexico (Montemayor, 1993).

Among the new generation of bilingual teachers of the region, an emerging consensus on the necessity of introducing literacy instruction in Na'huatl in the elementary grades poses a series of new pedagogical challenges on the level of school-wide language policy, classroom materials, bilingual instructional models, and assessment procedures. By necessity, local teachers would need to design their own reading evaluation instruments to begin the process of even conceptualizing the tasks before them that teaching literacy in an indigenous language implies.

As is customary in virtually all public elementary schools in Mexico that serve indigenous bilingual children, initial literacy instruction is provided exclusively in Spanish. Exceptions include pilot programs which have attempted to implement the long-standing Secretariat of Education language policy (DGEI 1990) to provide reading materials and learning activities in the students' indigenous language. The local school in which the present study was carried out falls into this category; however, literacy teaching continues, with occasional digression, to be in Spanish, the second language of most entering first grade children. Nevertheless, the new reorganized bilingual program represents a significant departure from previous practice; today, Na'huatl-speaking students are free to use both languages both on the playground and in the classroom, even to address their teacher in Na'huatl, an inconceivable occurrence under the former Spanish-only system (although this marked use of the indigenous language is still not common). Furthermore, official school policy emphasizes the recognition of the indigenous language (spoken by over 90% of the town's population) in various concrete ways: singing of the national anthem in both languages, introducing the Na'huatl alphabet, occasional writing, poetry, and declamation contests, and explicit teacher
Applications of cloze

Admonitions to value and preserve the community’s linguistic heritage (the majority of the school staff, including both administrators, is bilingual as well).

Thus fortuitously, an important convergence of interests is posed with reading teachers working with the languages of wider communication, who over the past decade, in turn, have been exploring classroom-based alternatives to standardized testing (Valencia, Hiebert and Afflerbach, 1994; Hodges, 1997). Our own experience with the cloze assessment sessions themselves, and subsequent initial analyses of the Na’huatl language cloze sample have confirmed our working hypothesis that, with appropriate modifications, the universally recognized usability and practicability of the technique can be matched by satisfactory levels of reliability and validity.

RESEARCH ON CLOZE AND LANGUAGE ASSESSMENT

Oller and Jonz (1994), in their recent volume, equate “meaningfulness” with “coherence,” tracing the origins of the cloze procedure to the early studies of the Gestalt psychologists. The ability to effect closure, combining fragments or isolated elements to form coherent and meaningful wholes, was considered, more generally, a measure of intelligence. Thus, early on, researchers proposed that the cloze procedure must be applied to self-contained, connected, and complete segments of discourse. In this respect alone cloze could be considered an integrative measure of reading; but in addition, the theory that cloze requires the simultaneous application of vocabulary knowledge, grammatical competence, sentence level decoding, and passage level comprehension strategies approximates the task to an even greater degree to the criterion of actual silent reading under normal, contextually constrained, conditions.

Particularly in L2 reading assessment, cloze appears to represent a valuable compliment, for example, to the analysis of oral miscues since the latter will tend to shift undue attention to the recoding aspects of oral reading. The second language reader, with only partial control over the L2 grammatical system, sound patterns, morphology, etc. will, by necessity, direct more attention to producing an oral version of the text that sounds more target-like, this, of course, at the expense of comprehension (Nurss and Hough, 1992). In other words, more conscious local processing takes limited working memory resources away from discourse-level processing. For a discussion of the relationship between oral output and decoding in the first language (L1) reading, see Salasoo (1986), and
Mandel-Glazer, Searfoss, and Gentile (1988), also McLeod and McLaughlin (1986) on controlled and automatic processing in L2 readers, and Zhang's (1988) study of Chinese ESL students' oral reading miscue patterns. Hypothetically, for the L2 reader, in "alleviating" the pressure to devote limited controlled processing resources to recoding surface features, cloze should provide for a more authentic measure. But as Miller and Smith's findings indicate, the effects of overt verbalization for readers of different ability levels are complex, and further investigation will need to examine these factors in both L1 and L2 readers.

Since the analysis of oral reading miscues and cloze testing share both integrative and on-line features, and both allow for qualitative scoring, they can perhaps be combined in a way that compensates for their respective deficiencies; cloze not being without its critics, who surely would point to the contrived and "unnatural" conditions that a fragmented text imposes, for example, on the inexperienced reader/test-taker.

On the related issue of cloze task sensitivity to sentence-level vs. text-level context, the research findings are inconclusive. Fixed-ratio deletion passes, at least, appear to be poor measure of awareness of larger context, with only a small percentage of items dependent on information across sentence boundaries (Cohen, 1994, p. 237). Bachman (1985), however, has demonstrated that the use of rational deletion criteria can significantly increase cross-sentence context dependence from the 10% that would by typical of a 1/11 fixed-ratio deletion. For beginner L2 readers, this modification would be especially effective since shifting the deletion method away from a random selection, toward, for example, more content words, will produce cloze passages that are more predictable in general, and, at the same time, more sensitive to discourse-level constraints (DeSanti, 1989). Also see Chavez-Oller et al., (1994), and for L2 readers, Hanania and Shikhani (1986).

Research on the critical issues in the assessment of second language learners' reading proficiency is part of a broader discussion on language assessment in general. In the literature, integrative approaches have come to be contrasted with what have been categorized as "modular perspectives," unfortunately, in terms that have often obscured the debate. The latter have been associated with discrete point methodologies and an analytic focus on structural units of language, the former, with "descriptive" approaches emphasizing holistic or global processing, and contextualized communicative proficiency closely tied to general cognitive abilities. Especially in the area of literacy assessment, the distinction is useful; however, reductionist models of language fall into the error of
counter-posing modular and integrative across all aspects of linguistic competence and language use proficiency. For example, strong versions of the "synergistic perspective" reject the autonomy of linguistic knowledge implying that all language processing is contextually constrained, meaning-based, and integrative in the most complete sense (e.g., corresponding to a kind of general system where a global/unitary factor governs the greater part of all linguistic and intellectual performance). Damico (1991), for example, argues that:

The internal structure of language proficiency is also integrated. The components of language (e.g., syntax, phonology, semantics, pragmatics, reception, and expression) are essentially terminological distinctions created in the mind for ease of discussion and analysis. They are not divisible and discrete in their functioning; they function holistically. Communication can only be assessed directly as it functions in naturalistic contexts, thus insuring linguistic reality (p. 178).

While in the area of classroom-based reading instruction and assessment, the notion of integrative and holistic processing is very attractive (and as we shall see, for valid reasons), the lack of theoretical clarity inevitably leads to serious practical problems. Evidence from extensive evaluations of bilingual students' language and literary skills has revealed important distinctions (linguistically and psychologically real) among the different components of school-related second language proficiency (Harley et al., 1990). Many aspects of language use in school are indeed closely tied to the general cognitive functions. This type of integrative processing, in fact, is characteristic of academic discourse, in its various forms, both oral and written. However others are not, or certainly less so, depending on the circumstances: for example, those aspects of language proficiency related to competence in the language code itself — the strictly linguistic aspects that in L2 learners often evidence remarkably uneven rates of acquisition. It appears that grammatical and phonological processing, for example, are more "encapsulated," autonomous, or modular — the aspects of language that some researchers refer to as the fast processes, that are "rapid and mandatory" (Sharwood-Smith, 1991).

On the practical level, confusing the modular and integrative aspects of language proficiency has led to egregious misinterpretations of assessment data in bilingual settings. The failure to separate the issue of linguistic competence from academic/literacy-related language proficiency has led to categorizing large numbers of bilingual students as
"semilingual" or "alingual" — in principle, a highly unlikely possibility for any language minority community, since such severe language disorders should represent a distinctly low incidence condition among all school-age populations.

In reading assessment, since the classroom teacher's concern will emphasize comprehension, it will be the discourse-level processes that must figure most prominently in evaluation tasks, precisely those that depend upon, to a greater degree, the intervention of top-down strategies and integrative mechanisms for constructing meaning. Recent trends in literacy assessment have de-emphasized the discrete point methods that sought to examine, analytically, the different components of decoding, again for sound pedagogical reasons. However, from the (correct) priority placed on evaluating discourse-level strategies and text comprehension it does not follow that all processing of written language is global, holistic, and integrative to the same degree.

Sharwood-Smith (1991), emphasizes that even if we accept a differentiated, modular-type, set of mental representations for the different aspects of language proficiency, this still leaves open the question of how language, in real time, naturalistic/authentic, uses (such as reading comprehension) is actually processed. In other words, the autonomous/interdependent knowledge structures that correspond to the different linguistic subsystems may, depending on the type of task, for example, interact in different ways, some more integratively, others in a more linear or encapsulated way. Reading specialists are familiar with the general concept of modularity from the implicit assumptions that underlie Oral Miscue Analysis (Goodman, 1994); subsystems of language as they come on line in the reading process interact. This implies that syntactic knowledge, semantic knowledge, phonological knowledge, etc. maintain a certain autonomy, to some degree, to be able to interact in the first place, a feature of language processing, for example in reading assessment, that allows for the analytical comparison of Observed response and Expected response. For example, in skilled reading, the linguistic subsystems are processed in an "integrated," and very efficient manner; in less-skilled reading often one or another subsystem predominates at the expense of the others.

The research on second language readers has perhaps contributed most to refining an interactive model where the role of discrete point, bottom-up, processes have received a more balanced treatment. Studying L2 readers text processing strategies under conditions of wide ranging variations in the level of L2 language acquisition (an opportunity not available in the case of L1 readers) has lent support to a version of the
Applications of cloze

modular perspective — a relationship of autonomy and interdependence among the language learner’s faculties that account for text comprehension. In studying L2 reading, less so in the classroom context, researchers often arrive at significant and important findings when discrete components are isolated for analytical purposes, again confirming that the global phenomenon that we observe does not correspond to an undifferentiated and completely context-dependent general purpose processor.

As the previous discussion of the cloze procedure indicated, this particular assessment tool would fall toward the mid-range on the continuum between fully integrative and purely discrete point assessments. Depending on a wide array of both modifications in the presentation of tasks, and interpretative approaches, cloze will tap the reader’s repertoire of strategies for utilizing context at different levels. If meaning depends on mentally constructing text coherence, the task of restoring a fragmented, “incoherent,” reading passage will call upon this set of discourse competencies at least to some significant extent. Local closure and parsing ability will also be revealed in respondents’ choices. Deletion procedures and scoring methods will determine, to a large extent, how sensitive the test actually is to text-level coherence and semantic and syntactic constraints at the sentence-level (see Cohen, 1994, on Structured-response formats, pp. 233-242). Particularly with L2 readers, a focus on both levels, separately and in regard to how they interact, will offer fruitful avenues of investigation.

SUBJECTS

To be able to examine students’ reading skills in Na’huatl, two grade levels were chosen that would hypothetically reflect growth in literacy skills (acquired “in Spanish” and available during the experimental situation, reading a text in Na’huatl); and with both groups beyond the initial literacy acquisition stage, relatively few “non-readers” would be participating in the study. Subjects included all 3rd and 5th grade students (four classes) in attendance on the day scheduled for each testing session (50 third graders, 54 fifth graders, for a total of 104 subjects). Including the training and practice items, sessions averaged one hour and ten minutes (time elapsed from initial instructions to the group to completion of the cloze passages by all subjects). Teachers’ reports confirmed that all students possess at least passive oral competence in Spanish and Na’huatl. Again, all 104 students learned to read and write in Spanish; experiences with Na’huatl texts of any kind had been incidental or sporadic at best.
DESIGN AND PROCEDURES

To achieve a measure of uniformity regarding access to previous knowledge schemata (both content and genre related) a traditional narrative of conventional, logical/causal, structure was selected. The story, "Soatl iuan miko" [The woman and the monkey] related a familiar theme of a young woman who bears a child after being kidnapped by an animal, escapes, and returns to her village where she and her son confront the injustices of society.

The deletion procedure sought to maximize context support for the task by: 1) maintaining a low ratio and providing a long lead-in; since the narrative was short, out of 339 running words, subjects completed only 15 blanks; 2) deleting only common, relatively high-frequency, content words (10 nouns, 5 verbs), in all cases, the final word in a sentence or phrase. While the omission of content words in final position ensures optimal textual predictability, careful omission procedures, even involving other grammatical functions, with sufficient sentence-level lead-ins, will provide adequate context support. Clearly, for more advanced readers, with more experience with Na'huatl texts, the above modifications become less critical. To facilitate the task, especially for young students who, faced with the cognitive demands of this kind of decontextualized reading activity, tend to guess at random, simply selecting words from the text itself, a list of the 15 exact words deleted, plus 10 distracters was displayed. Instructions indicated that words could be chosen either from the list of options, or subjects could provide their own responses. See Guthrie et al., (1974) for a discussion of the "maze procedure," a kind of multiple choice cloze format, similar in some ways to the modification in this study. Depending on the selection of each set of distracters, "maze" would lend itself particularly well to exact-word scoring (e.g., full credit for sentence and text-level compatibility, partial credit for sentence level compatibility alone, and so forth).

Since none of the students had ever participated in a cloze assessment activity of this nature, a brief training session was conducted with sample passages to demonstrate basic cloze test strategies. Before beginning the task, an overview of the story was provided orally, in Spanish, introducing the characters and the setting, and providing a general predictive framework for the event sequence, upon which students, with pencils in the desk, were asked to read through, silently, the deleted passage from beginning to end.
Scoring followed a rational cloze, qualitative rating procedure similar to the method described in MacLean and d'Anglejan (1986); also see DeSanti’s (1989) extension of cloze that incorporates coding categories from the Reading Miscue Inventory. A seven point scale differentiated among the following types of responses:

6-exact replacement or response that is syntactically compatible with the entire sentence and semantically compatible with the text as a whole;

5-the above criteria are met but response requires “some minor syntactic adjustment” (MacLean and D’Anglejan 1986, p. 819);

4-response is syntactically and semantically compatible at the sentence-level only;

3-the criteria in #4 are met, but response requires minor syntactic adjustment;

2-the reader’s response creates a sentence fragment (partial compatibility);

1-response does not form part of a grammatical fragment, indicates only a semantic relation to some reference in the text;

0-random response, not compatible syntactically or semantically with any fragment, no relation to any reference in the text.

In this study, subjects completed one cloze test as part of a larger investigation of biliteracy and language development. The same group of 3rd and 5th graders also participated in assessments of written expression in Spanish and in Na’huatl, the results of which are the subject of a separate report. Previously, a cohort of 2nd, 4th, and 6th grade bilinguals from the same school completed a battery of literacy assessments including Oral Reading Miscue Analysis, a cloze test, Oral Narrative and Writing. One of the methodological objectives of the investigation is to compare findings from the earlier, more individually oriented, evaluations of reading and writing to the present large scale group applications. If results from the latter evidence similar tendencies, this finding would lend support to the development of practical, classroom-based, instruments that are in fact accessible to teachers working in similar bilingual settings. However, the limited sample, methodological shortcomings, and the resulting tentative interpretations that are outlined in the next section indicate the need for further research on the application of cloze to the special circumstances of literacy development occasioned by the contact between a national language and indigenous language.
RESULTS

In this preliminary and exploratory phase of the study, it was important to determine if, under the conditions outlined above, actual student performance on the test was at least minimally consistent and reasonably stable. Subsequent analyses and comparisons, for example, with parallel cloze tests in Spanish, and writing samples in both languages, would depend on responses as a whole being non-random.

A comparison between the grades showed that 5th graders did out perform 3rd graders, as would be expected (but under the circumstances, not guaranteed) if students were responding meaningfully to the text. Converted to percentages, the mean score for 3rd grade was 31.8, for 5th grade, 45.2. A one factor ANOVA was performed with the Scheffe F indicating a significant difference between the grades; F(1,102) = 29, p<.005. For the entire sample, a split-half reliability check yielded a Spearman-Brown corrected r=.63. Considering that the number of deletions was far below the recommended 25 to 30 for maximum reliability (Bachman, 1985), and that for many of the younger students task difficulty resulted in a large number of responses that involved guessing (split-half reliability for 5th grade, alone, improved to r=.69), internal consistency could be considered adequate.

Given the special circumstances of reading in Na'huatl, and in particular, performance on an assessment instrument (for students, both a novel experience in general, and involving the application of skills in a language not normally associated with the academic setting), the dependability of the sample of tasks would be an important concern. Generalized confusion, rejection of the instrument, and the tendency toward random response would call into question the adequacy of the cloze items in regard to the proficiency being measured. See the discussion on estimates of internal consistency in Linn and Gronlund, (1995, pp. 81-90). In this regard, note the distribution of scores for individual cloze insertions, comparing third and fifth grade responses (Table 1).

Exact-word scoring, while generally yielding low overall ratings, would offer the classroom teacher a relatively rapid estimate for group comparisons and assessment of the appropriateness of classroom texts, for example. For the purpose of studying students' reading strategies however, qualitative scoring offers a wide range of analytical opportunities. For example, as part of a broader, on-going examination of interlinguistic transfer, cloze responses that were made in Spanish were coded separately. Comparing 3rd and 5th graders, a number of observations were noted that could provide data for further study.
Table 1

Distribution of Scores for Individual Cloze Insertions

<table>
<thead>
<tr>
<th>Score</th>
<th>Third Grade</th>
<th>Fifth Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>264</td>
<td>173</td>
</tr>
<tr>
<td>1</td>
<td>80</td>
<td>102</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>106</td>
</tr>
<tr>
<td>3</td>
<td>18</td>
<td>21</td>
</tr>
<tr>
<td>4</td>
<td>35</td>
<td>145</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>6</td>
<td>73</td>
<td>164</td>
</tr>
</tbody>
</table>

While it appears that 3rd graders tend to switch to Spanish with greater frequency (29 total switches/50 3rd grade subjects = .58, as compared to 20/54 fifth graders for a frequency of .38) the 5th grade students' use of Spanish material appeared to be more deliberate and purposeful. In general, we would expect compatibility ratings to be higher for Spanish switches since the reader obviously cannot be selecting at random, and would tend to be more aware of word choice as a result of a more conscious lexical search. In fact, the average rating for 3rd graders was 44.8 (13 points higher than their overall mean); for 5th graders compatibility increased to 64.1. Indeed, 5th graders may seem to exhibit a more selective switching pattern, possibly benefiting, to a greater extent (an increase of almost 19 points over their overall mean), from the opportunity to manipulate the two language codes in the written modality. The relatively low number of total switches (49 out of 1299 total responses) advises caution against any early interpretation; however comparisons with similar code alternation phenomena in writing and oral discourse may reveal significant parallels. In any case, the switches themselves (together with the number of new words supplied from students' own lexical searches, plus alternate spellings of choice words) indicate that in these cases response patterns were far from haphazard. The examples shown on Table 2 are representative of levels 3 & 4 and 5 & 6.

The results below on interlinguistic switching are offered as one possible avenue for further research that has received little attention in the field of second language reading. Cloze in particular appears to present special advantages in tapping this processing phenomenon that is the object of widespread interest among teachers working in bilingual schools.
Table 2

Cloze Responses with Spanish Switches (in italics)

Level 3: sentence level compatibility requiring minor grammatical adjustment
Level 4: sentence level compatibility
Level 6: text level compatibility

Hector — 5th grade
Oasito kuajtlan, omajsito inauak in xolopijtli uan okimakak in chango. (level 4) [He came to the forest, he came to where the devil was and he gave him the monkey]

Elena — 3rd grade
Nikojkoltsin oknikia momachtis, oktitlan inauak se inicuela. (level 3) [His grandfather wanted him to learn, he sent him to a the school (exaxt word answer: priest)]

Petra — 3rd grade
Opank miek tonaltin, in pipiltontsin omochi telpokua uan sepa kuak amo ompa okatka in miko, otlatlapo okimemej ni nantsin uan okuikak ompa ni poehlo. (level 6) [Many days passed, the boy grew to become a young man and once when the monkey wasn’t there he opened (the cave) and carried his mother away and took her to the town]

Inocencio — 5th grade
Ompa nijki amo okxiko in teopixki, okilnamik “Nikitlanis kuajtlan una ompa mak kuakan in tlacatl; (level 4 Na’huatl response) uan niktlaJKuiluilis in xolopijtli kampa kampa makiku, una ijkJon tiyolpachiuske uan ompa lamis nin pleito.” (level 6) [There the priest couldn’t stand him either, he thought “I’ll send him to the forest so that the man (exact word answer: animals) eat him up; and I’m going to write to the devil where he lives so that he eats him up, so that we can live in peace and that way will end this dispute, quarrel]

DISCUSSION: RESEARCH DIRECTIONS

Regarding the applications of cloze, two objectives were outlined in the introduction to this report: 1) to explore some of the features of the technique among its numerous variants and modifications as an investigative tool; 2) mindful of the time and resource limitations of the
classroom teacher, to propose its expanded application as a teaching approach. In the area of literacy, at least, productive research models often directly inform sound classroom practices; see for example the examination of integrative and discrete point perspectives in the above section: "Research on cloze and language assessment." Thus, our discussion of the study will begin with (1), and conclude with some general considerations regarding (2).

For an initial probe regarding the suitability of cloze testing into the relatively uncharted domain of literacy assessment in vernacular languages, the results seem to offer a well-grounded starting point. The challenge will be to improve reliability while, at the same time, maintaining high levels of redundancy and context support. As was evident from the relatively low success rates, increased redundancy and context support represent a minimal condition in situations of incipient or initial literacy development; in this case, literacy development extended to a previously "non-academic" language. Under the circumstances, the following adaptations would contribute to greater consistency: 1) with increased exposure to, and practice with, the activity, time devoted to training sessions can be applied to the task itself — longer reading passages resulting in more deletions (We are still hesitant, however, to increase the deletion ratio much beyond 1/20. Moreover, interest in the activity and general attentiveness were actually significantly underestimated, suggesting that students would easily tolerate longer sessions); 2) preparing texts of varying difficulty levels to accommodate both below grade level readers, as well as Spanish dominant passive bilinguals with a tenuous command of Na’huatl.

Regarding considerations of validity, improving the "authenticity" of the assessment task itself: 1) greater context support previous to the actual cloze activity, including a more complete activation of story and content schemata, accompanied by a graphic representation of the narrative, would provide needed extratextual redundancy for many readers; 2) directions and previous knowledge activation should be presented in the language of the text, instead of Spanish. In this case, increased context support should result in a greater dispersion of scores, in turn, actually favoring reliability.

The superior performance of 5th graders over 3rd graders would hardly deserve even passing mention if the cloze test were in Spanish (the language of virtually all literacy instruction in this case). While the results of the Na’huatl cloze were to be expected, the outcome could conceivably have been different, a smaller, non-significant, difference for example. Given the acute sociolinguistic imbalances between Spanish
and Na'huatl, bilingual students performance on language tasks in their indigenous language, in a school setting, are often significantly affected (e.g., negation of linguistic knowledge or refusal to exhibit competence in the latter). In addition, despite the minimal presence of Na'huatl in daily academic discourse, the apparent access to comprehension strategies appears to confirm models of interdependence (Cummins, 1994) based on research in more "balanced" bilingual contexts.

Regarding evaluation of reader insertions, qualitative approaches are clearly preferable in the case of student/student peer correction sessions. On the other hand, for certain limited purposes, exact-word evaluation can provide classroom teachers with useful information. In this regard, an interesting comparison is found between the quantitative exact-word scoring results and the seven point scale. The exact-word method yielded decidedly low scores for both groups: 8.8% and 16% respectively. However, despite the narrow spread of scores, concentrated at the low range, the correlation with the relative compatibility scores (r=.78, p<.005) suggests that for classroom purposes the exact-word method could be useful to teachers for global, whole-class measures, for example, of text difficulty of reading material in Na'huatl. Evaluating individual responses for acceptable substitutions for the exact word and partial credit is entirely too tedious and time-consuming for most classroom teachers.

The opportunity for examining both interlinguistic transfer and access to discourse processing skills is far from exhausted by the precursory analysis of Spanish switches reported above. In contrast, again, to oral miscue analysis, the processing of cloze passages is less "data-driven," (in this regard resulting in a less "authentic" task), allowing for greater "on-line" reflection, e.g., involving language choice decisions. Our own findings from the previous study of 2nd, 4th and 6th grade readers found cross-linguistic switching in oral reading to be extremely rare. We are reminded, here, of Hanania and Shikhani's (1986) findings that cloze tests and writing assessments showed a closer correlation than initially expected. Cloze tests, especially under untimed conditions, should allow for greater degrees of attention to both structural features and meaning relationships within and across the text, as in written expression.

Previous analysis of our 2nd, 4th and 6th grade cohort's writing sample (Francis, 1998) suggested an interdependent relationship between the development of children's metalinguistic awareness and patterns of lexical borrowing from Spanish when writing in Na'huatl. As may be indicated, tentatively, in our 5th grade cloze readers' Spanish switches, with experience in manipulating the bilingual system, older students use their
translinguistic resources in a more reflective and meditated way (of the 29 Spanish switches made by 3rd graders, 7 received compatibility scores of 0, while only one 5th grade switch resulted in a completely incompatible response). Again, it is important to emphasize the speculative nature of any interpretations of subjects’ switching patterns on cloze responses, primarily in consideration of their low frequency. The tentative analysis, rather, suggests a line of future research into this important aspect of bilingual literacy development. However, keeping the tentative nature of the analysis in mind, the apparent tendency on the part of 5th graders to apply this strategy in a more conscious manner is entirely consistent with previous findings (Francis, 1997). Older, more mature, bilingual writers (6th graders in this case) showed a tendency toward avoiding content word switches from Spanish to Na’huatl; and on assessments of basic vocabulary knowledge they showed the same tendency (e.g., to substitute the “borrowed” Spanish items with Na’huatl equivalents.

Orthographic choices, “interference” features in reading and writing, and the transfer of grammatical patterns are all strong candidates for further research, with important pedagogical implications for teaching literacy in bilingual contexts. Among both teachers and parents in language minority communities, these issues are the source of persistent, on-going, discussions and controversies.

Related to our discussion of modularity in general, future research can examine how some aspects of bilingual children’s language proficiency may show developmental trends that differ from others. For example, findings from previous phases of the present study suggested that the advanced that children demonstrate in constructing global coherence mark a steady upward trend, while the lack of control over certain surface features (related to interlinguistic transfer) seem to be almost impervious to correction. On the other hand, we can take note of developing maturity in regard to switching patterns that are evidence of bilinguals’ enhanced control over other aspects of transfer.

The application of comprehension skills to reading in Na’huatl is a complex phenomenon. Comparing students’ Spanish cloze tests will possibly shed new light on this promising research domain as well. In any case, it would seem reasonable to include both discrete-point type analyses (that correspond to the modular aspects of linguistic transfer), and more integrative measures (that can account for the top-down functions) in evaluating and studying second language literacy in all its varied and multifaceted aspects. Appropriate to our theme, Kalantzis, et. al., (1990, p. 206) point out that, in practice, the discrete-point/integrative
distinction coincides with the dilemma of the "conflicting demands...of validity and reliability." Indeed, the respective clusterings of modular/bottom-up/discrete-point/reliability versus unitary-global/top-down/integrative/validity reflect complementary perspectives on the functioning of very complex interacting systems.

CONCLUSION: PEDAGOGICAL APPLICATIONS

The application of cloze to language and literacy teaching seems to have received relatively less attention (see Greenewald, 1981; Jacobson, 1990). If in the domain of assessment applications the research debate has yet to produce a broad consensus, in the area of teaching literacy skills the advantages of cloze are less controversial; they reside primarily in its wide-ranging elasticity. An assessment technique with potentially high content validity, in fact, should lend itself well to instructional applications. For example, an important feature of informal classroom-based evaluation is maintaining the option to be able to shift dynamically between teaching and testing, as well as completely integrating, when appropriate, the two pedagogical activities. Depending on the deletion method, readers' attention can be drawn to specific linguistic features that are the object of instruction, particularly, structures that L2 learners appear to be ready to master.

In this regard, since the procedure can be readily modified to maximize natural, text-based, redundancy, one of its obvious advantages is that teachers with minimal levels of proficiency in their students' L1 vernacular, for example, can with great facility generate an entire series of reading comprehension activities in the language. This particular benefit of cloze is especially relevant to the majority of school settings where in fact teachers do not speak or understand the indigenous language of the community. Since summative evaluation is not the objective of the activity, students who are speakers of the language evaluate their own and their peers responses. Small groups, formed heterogeneously, or by language ability, according to the instructional objective of the day (among other learner-related considerations), effect closure of the texts through the invaluable process of reflecting upon, confronting, and debating linguistic options. Here, qualitative rating scales, such as those referred to in previous sections, provide the optimum framework for developing language awareness and comprehension monitoring. See Block
Applications of cloze  41

(1992), for example, on developing “metalinguistic control” in L2 reading.

Here, in fact, is where assessment and learning tasks are most productively integrated; self-referenced evaluation, peer feedback, and interactive reflection promote a more analytical posture toward language and language learning (Genesee and Upshur, 1996). Cloze-pasages focus learner’s attention on relevant features of the text, and provide the opportunity to view the task of comprehension from the point of view of problem solving. For practical suggestions on the use of cloze in L2 teaching, see Wallace (1992) and Nuttall (1996).

In teaching and in assessment, the same fundamental principles outlined above apply. The selecting of learning tasks, however, shifts the focus toward authentic and representative samples of language — a kind of content validity consideration. Depending on the deletion choices made in designing cloze passages, the classroom teacher can alternately focus on one or another comprehension monitoring strategy, at whatever level of decoding meaning that seems to present special difficulties for the L2 learner. Interactive models of second language reading (Nurss and Hough, 1992) provide for the flexibility that teachers require in deciding to which aspects of top-down and/or bottom-up processing learners’ developing metacognitive strategies should be directed as they gain proficiency in both literacy and their second language.

More generally, the development, at the community and school level, of literacy learning materials in the indigenous language has become a vital necessity given the limited production of vernacular language primers and grade school readers available from the Secretariat of Public Instruction in Mexico City. In addition, informal classroom assessment of reading in both Spanish and Na’huatl will offer bilingual teachers new insights into the rich and complex interaction among the different aspects of their students’ language proficiency. In this respect, researchers’ theoretical interest in probing this new field of study coincides with local educators’ search for new directions in second language literacy teaching.

NOTES

1. In regard to the comparison between oral miscue analysis and cloze testing, the tension between validity and reliability is a pertinent consideration. Unlike in other domains, language assessment is faced with a kind of partial trade-off that places these two requirements of assessment into a kind of contention of sorts. Maximizing reliability is
often achieved at the expense of validity and vice versa (Alderson et al., 1995, pp. 186-188). Davies (1990) frames the dilemma in terms of “being explicit as to what is being tested... and controlling uncertainty through statistical operations” (p. 53). In cloze testing, for example, a decrease in the frequency of deletion provides for more redundancy, consequently favoring validity, but possibly at the expense of internal consistency of responses. However, since cloze is more “test-like,” and potentially amenable to manipulation that will improve reliability, it can serve as a useful compensator for, or corrective to, more “authentic,” but less controlled (or controllable) reading assessments.

2. To what degree cloze passages are sensitive to discourse-level constraints remains one of the central controversies in the field. Citing studies by Kamil et al. (1986), Bernhardt (1991) argues that attention to sentence-level constraints is sufficient for correct responses. Also, see the discussion in Storey (1997). For example, subjects’ scores were not significantly different between deleted texts, where sentences were randomly rearranged, versus texts with normal, coherent, sequences. Bernhardt’s strong critique of cloze, however, seems directed toward traditional designs (fixed ratio deletion and exact word scoring). In this regard she correctly observes that for a L2 learner, often “cloze testing is a vocabulary exercise, not an assessment of reading” (p. 197).

Further research could focus explicitly on this interesting research question. For example, applying a series of rational deletion designs, the degree of discourse-level compatibility of responses can be examined using MacLean and D’Anglejan’s (1986) qualitative scale, that specifically takes text-level constraints into account. A series of experimental cloze passages could include the following: 1) deletions in which, in fact, only sentence and clause-level clues are necessary for closure; 2) a finer distinction could be highlighted by differentiating between responses that require semantic or syntactic processing alone; 3) deletions that explicitly require the application of discourse-level comprehension strategies for “full credit,” as opposed to responses that are compatible syntactically and/or semantically at the sentence-level only.

Bernhardt proposes the recall protocol as an alternative, a technique that has proven to be reliable if scoring templates are carefully designed to insure interrater consistency. However, for the purposes of the present study recall protocols present two limitations: 1) if retelling is oral, administration of the assessment must be individual; 2) if recalls are solicited in writing, the intervening factor of proficiency in written composition skills is not possible to control, especially in the case of younger children.
3. The role of context, first in reading acquisition and later in proficient reading has been the subject of considerable discussion. Specifically in regard to cloze, it is important to distinguish between two broad categories of context. Oiler (1992) refers to the principle of scaffolding in initial literacy learning: how “supplemental context” in situationally dependent reading activities supports the task of “linking linguistic forms with facts of experience” (p. 61). Here, the novice reader constructs meaning from written language input where the “fit” between text, previous knowledge and even situational context is, optimally, complete.

Depending on appropriate modifications and its form of presentation, cloze tasks begin to require the beginning and intermediate reader to shift away, strategically and reflectively, from extratextual context toward the use of the particular kind of context that characterizes school literacy — textual referents alone, situated inside the text itself. See Olson (1994) on the “discourse of literal meanings” where texts are taken as “closed.” Developmentally, a shift occurs: from the manipulation of situations and concrete categories to the mental manipulation of their representations.

4. Completely avoiding function word deletions also seemed to provide the opportunity to be able to differentiate, to a greater degree, between use of text-level compatibility and local, sentence-level context use, respectively, Bachman’s (1985) type 3 and 2 levels of context (p. 539); in our schema, categories 5/6 and 4/3 (see table 2).

REFERENCES


Norbert Francis is a faculty member in the Department of Education at Northern Arizona University, in Flagstaff.

The author gratefully acknowledges the support provided to the project from the Office of Grants and Contracts of Northern Arizona University and the US/Mexico Fund for Culture, and to Pablo Rogelio Navarrete Gom’ez for his assistance in the analysis of Na’huatl language texts.