

Awareness of Form and Pattern in Literacy Assessment: Classroom Applications for First and Second Language

Silvia-Maria Chireac

Universitat de València

Norbert Francis

Northern Arizona University

John McClure

Northern Arizona University

ABSTRACT

The study examines performance on an evaluation instrument for classroom use for the purpose of improving instruction that focuses on awareness of language in reading and writing. Findings are reported on the piloting of the first in a series of literacy assessments with students from a rural elementary school in Ecuador. All students who participated in the pilot study are first language speakers of the language of instruction, Spanish. The report follows up on two recent studies on: awareness of language in self-correction (writing ability of bilingual and second language learners, Mexico), and overall literacy attainment comparing boys and girls in rural schools in Latin America with the objective of confirming or disconfirming recent UNESCO estimates of gender parity. The authors present a proposal for monitoring literacy learning in elementary school that will help teachers to integrate teaching and evaluation, an approach that can more effectively link classroom testing to actual instructional practice and student learning outcomes. Further development of the materials presented for discussion can be especially useful for monitoring progress in second language learning and bilingual development.

Key terms: formative assessment, summative assessment, language awareness, literacy learning, Spanish

INTRODUCTION: A CONCEPTUAL FRAMEWORK FOR THE EVALUATION OF READING AND WRITING

How educators understand literacy learning determines the methods we use for evaluating the progress that our students make in becoming proficient in reading and writing, in particular reading and writing for academic purposes (Rogier, 2014). The approach that we favor in the current project is one that attempts to estimate this proficiency at the levels of orthographic knowledge, sentence processing and discourse comprehension/expression with the idea of better understanding the components of literacy. In this study, no proposal is put forward regarding the specificity, or non-specificity, of the components of reading and writing ability. The kind of assessment tools discussed in this report are not of the kind that can provide evidence about questions of specialization, domain specificity or domain generality of any one or another aspect of literacy knowledge or skill; see Foorman et al. (2011), Lonigan et al. (2013) and Shatil and

Share (2003) for findings and discussion. Thus, the idea of “components” should be taken in the more informal or broad sense, the sense that we have in mind when thinking about how students learn the skills and abilities related to word identification, spelling, sentence production, text comprehension, coherent expression in writing, and so forth. Research on how these skills and abilities develop has suggested that awareness, focused attention, is an important aspect of ultimate attainment when advanced comprehension and expression is important (Ehri, 2014; Gombert, 2005; Svalberg, 2007; Tong et al., 2013). Focused attention and awareness is one of the key dimensions that the analysis of the pilot cloze assessment sought to measure (see the Discussion section below regarding the relevant result).

For example, the foundation of fluent reading, necessary for advanced literacy in both first and second language, is the development of efficient word recognition skills (Frasier, 2004, Share and Stanovich, 1995). At the text level, effective comprehension depends on the ability to attend to and reflect upon ideas and concepts, and how they are organized into larger units (Samuels et al., 2005). In the same way, in skilled writing, an “analytical” awareness of language is necessary for on-going monitoring and self-correction even on the first draft attempt. How, then, can teachers measure progress in these (different, according to the hypothesis) aspects of literacy learning?

The other consideration related to the relationship between assessment and teaching is the direct positive effect that the testing procedures should have on student learning. Ideally, learners should benefit, directly, from working on the problems they are presented with in the assessments we give them. The problem-solving tasks include the training activities they work on in preparation for an assessment and the review of correct and incorrect responses after an assessment. This is one of the ideas behind the concept of formative evaluation. In turn, measuring progress, of the summative kind, should be useful for program evaluation: is the literacy teaching program effective? Answering this question then counts as an indirect, longer term, benefit to learners.

METHOD

Participants and community

The site of the study was the town of San Lucas (approximately 6,000 inhabitants) located on the Pan-American Highway, in northern Loja Province, Ecuador. A rural economy (predominantly cattle raising, cultivation of corn, potato and bean), 91% of residents self-identify as belonging to the Saraguro ethnic group, historically Quichua-speaking. In San Lucas, today, only a small percentage of the population speaks the language. Children enrolled in the public elementary school are all native speakers of the national language (NL) of instruction, Spanish. Knowledge of Quichua in the school-age population is rare. Consistent with the official Ministry of Education curriculum (DNEIB, 2012), students receive weekly instruction in Quichua as part of a language revitalization program. The performance of 41 sixth grade children (all 6th grade children enrolled in the school) on the pilot reading comprehension test was examined. They are part of a larger group of 95 students (including twenty 2nd graders and thirty-four 4th graders) who are currently participating in the pilot study.

Three assessments

Results from the piloting of the first assessment, of reading ability, will be reported below. In this section we discuss the design features of the other two tentative components of the literacy evaluation for the purpose of presenting an overall framework of how they might be related upon completing the design of all three components.

Reading – the cloze test

Of the different variants of the cloze procedure the project selected the text-level narrative type to tap into, potentially, comprehension processes beyond the sentence level in addition to word identification skills at the sentence level. For each of the 20 items, the reader selects the correct response from three choices (see “Historia de los niños que vivían en el bosque [Story of the children who lived in the forest]” Appendix 1). Omissions, together with the selection of distractors, were chosen primarily to require that the reader take subsequent sentence-level context into account (14 items). For six items, previous context provides sufficient information to eliminate the two distractors. Nevertheless, in all cases, all previous context from the beginning of the story might provide the reader with relevant information for a correct response. An omissions/total-word-count ratio of 1:9.6 was maintained by generally avoiding more than one omission in any independent clause, which was the case for 18 of 20 items. For example, in item #9 the omission appears in a dependent clause, in items #13 and #14 the omissions appear in two separate independent clauses (see Appendix 1).

Items #9 and #10:

Cuando completó _____ veces siete, se le cayeron las _____....
(dos, los, la) (manos, piernas, brujas)

When [she] completed _____ times seven, from her fell off [her] _____....
(two, the [masc. plural], the [fem. singular]) (hands, legs, witches)

Items #13 and #14:

Salió al _____, se quejó y empezó a _____.
(otro, árbol, patio) (volar, velar, valer)

[She] went out to the _____, [she] groaned and began _____.
(other, tree, patio) (to fly, to watch over, to cost)

Almost all omissions required a content word for completion. We are proposing that the present closed-ended (choice-question) design represents an improvement over the project’s previous more open-ended (limited-response type) application of the cloze procedure which provided a word bank for each page of text. The word bank consisted of a list of correct responses plus one distractor. In turn, our hypothesis is that the multiple-choice format might be superior overall, especially for beginning readers, to the traditional cloze test that provides no choices. The latter no-choice design appears to impose processing conditions that are simply too onerous for all except the most advanced readers, potentially leading to unreliable results, especially in the

lower grades and for beginning second language learners of the text language. Even using grade-level material, response patterns, depending on circumstances, can end up being massively random for a significant portion of the participants, in addition to results with many items left blank.

In a study by Mostow et al. (2017) the interesting design features of the cloze test that provides choices for each omission in a closed-ended design are discussed at length. For example, the selection of distractors allows for managing the task difficulty of items and possibly studying different aspects of word identification and comprehension processes. Among the three incorrect choices distractors can be systematically included that are:

- ungrammatical,
- grammatical but semantically incompatible at the sentence level, and
- grammatical and meaningful in the context of the sentence but incompatible at the text level.

The authors' design emphasized predictability even more than in our proposal by omitting only the final word of a sentence, with a resulting omissions/total words ratio that is even higher. The point is well taken, as ratios lower than 1:9, in our view, begin to undermine the very purpose of the assessment: estimating reading comprehension approximating actual text processing demands, an important consideration of validity. An interesting contrast to the Mostow study, together with the present closed-ended item proposal, is the discussion by Brown (2013) regarding the limitations of the traditional cloze procedure (no multiple-choice) in a wide-ranging retrospective of the research. A relevant exploratory study, parallel in a number of ways to the present study from Ecuador, applied the traditional cloze procedure with young readers: average age 9 years 5 months, students from São Paulo and Minas Gerais, Brazil (Oliveira et al., 2011). While an overall positive evaluation of the usability of the instrument was reported, we look forward to further analyses of the results regarding response rates and item difficulty.

Error correction

Separately, a series of three error correction evaluations was designed utilizing the same text to probe the initial acceptance on the part of students of the assessment tasks. Future piloting will select the one error correction test that yields the best performance across the grades to accompany the cloze test. The correction task requires students to read the story and identify as many errors, previously introduced into the text, as possible, and then supply the correct form. The assessments seek to estimate students' ability to identify and correct, on the three separate instruments, errors of spelling (20 errors total, Appendix 2), grammar (20 errors, Appendix 3) or punctuation/capitalization (25 errors, Appendix 3).

Target errors of orthography on the spelling test include 14 homophonous or near-homophonously spelled non-words, and 6 similarly sounding words or non-words. On the grammar correction test, 19 errors are correctly spelled words plus one omission (e.g., errors of concordance, subject-verb agreement and gender, all of which correspond to basic sentence-level grammar knowledge of native speaking children, not aspects of higher-order academic register). On the third correction test, all 21 capitalization errors occur in real words, otherwise, spelled correctly. All 4 target punctuation errors consist in the omission of a period (a clue is provided by inserting a space and correctly writing the first word of the next sentence with an upper-case letter). Thus, on each test, in turn, readers are specifically directed to attend to spelling errors, to

sentence grammar errors, or to grammar problems at the sentence level and beyond occasioned by the error of punctuation or upper/lower case.

Each item on the three instruments receives a maximum score of 2 (successful identification + correction), 1 (identification), or 0. We recommend that in the application of the cloze test-error correction test pair that the former precede the latter, preferably on separate days. During the application of the cloze test, students can be told that if they select all the correct words there will be no omissions or errors remaining. In any case, the tests should always be given in the same order from one group to another: cloze test first. Using the same instrument for 2nd, 4th and 6th grades, if the complete piloting project so indicates, direct group comparisons can be easily made. For obvious reasons, none of the proposed assessments are recommended for use as tests in 1st grade. For 1st grade, the materials can be used as a group learning activity, with the teacher reading the text aloud to the class.

This design, we propose, is an improvement over our previous (open-ended) design for error correction in which students freely selected which errors to correct on their own first draft compositions. On the previous design, after a training session that introduced the children to correcting and revising techniques, attempts were scored for effectiveness (degree to which the revision improved the original text segment) and for level of text (word, sentence, discourse). Aside from considerations of practicality (the test is usable only by full-time researchers), there is no way to systematically focus the assessment on learners' ability to attend to specific error patterns of interest to the teacher and the instructional program. Hypothetically, the potentially higher face/ecological validity of the error detection-and-correction test, as compared for example to the cloze, in its transparent connection to the all-important editing skills in writing (Ferris and Hedgecock, 2005) should favor consistent results.

Written expression

Finally, to complete the series, a writing sample was taken from students, utilizing a completely open-ended design, based on a narrative model presented orally with graphic context support ("El cazador de venados [The deer hunter]," Appendices 7–10) to gauge acceptability. Following up on the successful application of this prompt from the Mexican study (Francis, 2017), a satisfactory initial response (full participation by all grade levels) was received, pending analysis for coherence and inter-grade level comparison. As a typical classroom literacy activity, no previous training session is required, as was widely confirmed in the previous Mexican study in two separate second language learner communities, with participants starting at age eight. Nevertheless, providing a narrative model (serving as a story-grammar framework) from which to compose is an important feature to ensure consistent access to a previous knowledge schema for the writing task, necessary for comparison purposes. Foreign language and indigenous language versions of the writing assessment follow the same protocol with a different story of equivalent event structure and narrative pattern, again for comparison (see Francis, 2012, for implementation in two languages for bilingual or second language learners).

Procedures

Especially taking into account the relevant experience of the youngest beginning readers, evaluator and students work together on the cloze training activity ("El rey Midas [King Midas]" Appendix 5) to solve all of the items. Learning the logic of the cloze test, for example how to

process both previous and subsequent context, and in the present case, for understanding the concept of distractor, is important for maintaining comparability across the grade levels. The evaluator selects a distractor to prompt students: why the wrong choice produces an ungrammatical sentence or a sequence that is not semantically compatible with the rest of the sentence or with the larger passage (“doesn’t make sense”).

For the error detection and correction assessment, students will first work with the teacher/evaluator on the correction of the training texts (Appendix 6):

- how to be attentive, respectively, to errors of spelling, grammar or punctuation/capitalization, and
- reviewing the criteria for an acceptable attempt by demonstrating positive and negative examples.

After helping students make the appropriate revisions, in the same way as on the cloze training activity, the teacher/evaluator deliberately selects a non-target word (no error) to modify incorrectly. Students are prompted to explain how a correct form was made incorrect. We emphasize, that except for the writing sample for which no practice session is required, students should not attempt any of the assessments without first working with the teacher/evaluator on the corresponding training activity. After reviewing responses on the practice test as a group, students submit a completed attempt with responses to all items answered correctly. The first two assessments, cloze and error correction, should be given on different days, and students should not be allowed to consult their cloze test page while working on the error correction assessment.

RESULTS OF THE READING COMPREHENSION ASSESSMENT

At first glance, grade level tendencies in performance suggest that the cloze test is usable for the purposes outlined in the previous sections. Expected satisfactory results from grade 6 were projected as the target performance for elementary school graduates, reason for our focus on this group of students. Overall, the mean correct result for 6th grade of 78%, with four scores of 100% and 3 results out of 41 below 60%, with the lowest score of 50%, corresponded to expectations. For checking purposes, a random sample (N=20) taken from a secondary school in the same community, indicated a hypothetical upper range in performance, (mean—91%) approaching completeness among upper-grade students. Crucially, the differences in performance both between 2nd and 4th and between 4th and 6th grades are significant, respectively: $t=5.57$, $p<.001$, and $t=5.31$, $p<.001$. The progress of boys and girls across the elementary grades is comparable, the small divergence in 6th grade (1.34 points) not significant: $t=1.37$, $p=1.35$. The internal consistency of the 41 student responses was estimated as moderate. The calculation was based on 18 items of the total 20, given that two items yielded no errors: estimated reliability using the Cronbach alpha measure, $r=.686$. No item on any of the 41 test papers was left blank.

Figure 1. Reading comprehension: Cloze test

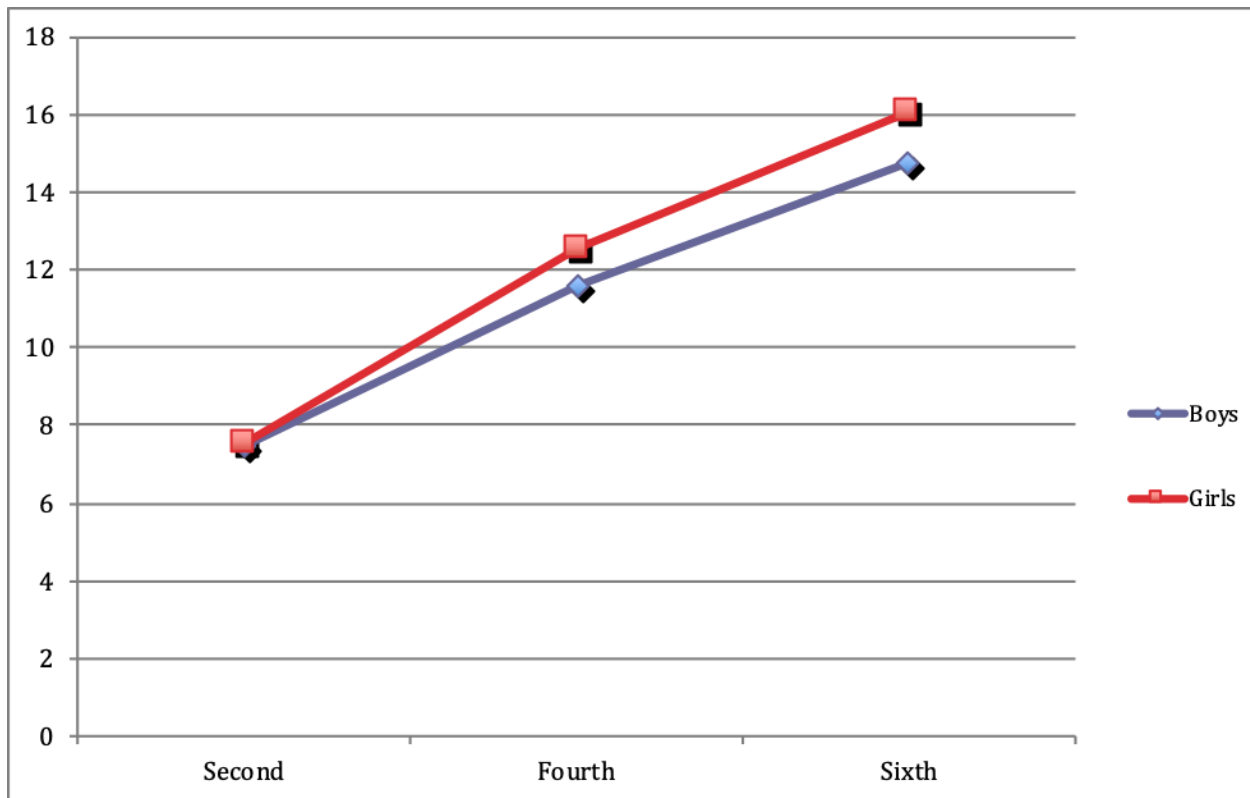


Table 1. Reading Comprehension: Mean and Standard Deviation for Grades 2, 4 and 6

	2 nd grade		4 th grade		6 th grade	
	n	M (SD)	n	M (SD)	n	M (SD)
Boys	11	7.45 (2.16)	17	11.59 (3.21)	11	14.73 (2.15)
Girls	9	7.56 (3.17)	17	12.53 (3.54)	30	16.07 (2.97)

Regarding the range of item difficulty, as mentioned above, for two items there were no errors (difficulty index of 1.000), with the most difficult item (#12) yielding an index of .342, in effect an outlier as the only result below 50%. Recall from the Methods section that for most items (14 of 20) the omission (correct response) and distractors were designed to prompt the reader take subsequent sentence-level context into account. For six items, previous context provided sufficient information to eliminate both distractors. Examining student responses along this dimension, we can tentatively point to a relationship between the requirement to process the entire sentence level context (implying a kind of “retrospection”) and difficulty. A significant Pearson Correlation of $r = -.51$, $p < .05$ resulted from comparing the difficulty index with a measure of “retrospection” – how much of the text beyond the omission (how many words after the blank)

need to be processed for the elimination of both distractors. The “retrospection” measure ranged from 0 (for the group of six “previous context” items) to 7.

DISCUSSION

On the cloze test, the correlation between the need to use subsequent sentence context and difficulty suggests that in the design of reading assessments this feature can be considered in presenting students with a broader range of different kinds of tasks and challenges, as recommended by Mostow et al. (2017), above. Hypothetically, the more difficult items required more deliberate attention to meaning and grammar patterns and higher degrees of controlled processing at the word and sentence level. Thus, this dimension, related to sentence complexity, can be varied to generate a range of non-trivial difficulty among items. The finding, by the way, coincides with previous results from our Mexican assessment project: students who were able to recover from reading errors (self-correcting) by successfully attending to wording *after* the error were shown to be, on average, more attentive readers and writers on other measures (Francis, 2012: 231—251). We attributed the development of this skill set to be related, hypothetically (subject to verification in future studies), to more advanced levels of language awareness (Chireac et al., 2019, in press). The construct underlying this dimension of item difficulty (in that it follows from general theoretical principles of information processing in reading), plus the finding of progressive and consistent advances across 2nd, 4th, 6th and secondary grades, we propose can be taken as evidence in favor of overall validity.

The present alternative for classroom literacy evaluation, in its three parts, described in the previous sections, proposes for future research the idea that language awareness forms part of the central core of advanced literacy ability. Except perhaps for the tentative suggestion related to the “retrospection” analysis on the cloze test, the findings of this partial evaluation of the testing instruments do not yet provide empirical support for the proposal. Rather, the different features of these assessments that are conceivably related to language awareness, or metalinguistic awareness, appear to be tied to the kinds of knowledge and skill that support advanced proficiency in academic uses of language (Leikin et al, 2005). They appear to be the kinds of abilities that teachers promote when students are challenged with reading and writing tasks that are more demanding. Metalinguistic awareness (MA), if the hypothesis is shown to be correct, presents itself at all levels of processing when learners go beyond basic implicit language ability (as in everyday conversation):

- beginning with phonemic and orthographic awareness, for young readers in particular,
- other kinds of focus-on-form attention at the word level,
- awareness of grammatical patterns as this awareness is used in constructing sentence meaning, and
- awareness of structure (involving concepts, events, etc.) in paragraph and text coherence building.

In the higher-order domains, MA can be understood as metacognition applied to discourse-level comprehension, reflection on and monitoring of understanding. Metacognition of this kind comes into play, for example, in the cognitive demands of challenging expository text reading in school (Yau, 2009). Metacognitively aware readers apply self-monitoring strategies such as detecting the breakdown of comprehension to then engage in self-correction (Samuels et al., 2005). When composing, writers would do the same when checking their own writing for correct spelling, sentence meaning and text-level coherence, all of this the hallmark of academic

language proficiency in one of its core acquisitions: secondary discourse ability (Bongartz, 2016). These kinds of ability were the ones that we had in mind in the design of the error correction tests in particular (Appendices 2, 3, 4).

In regard to the features of the cloze test, while frequent backtracking in reading has been shown to correlate with difficulties in decoding among poor readers, other aspects of retrospection (“looking and thinking back”) might be shown to be associated with monitoring strategies characteristic of skilled and efficient reading. Hypothetically, monitoring takes effect at the point of contact where accurate decoding and comprehension processes come together, in a kind of self-regulation. Monitoring can be thought of as real-time problem solving and “troubleshooting” applied to revision strategies, as in the case of the repair of comprehension breakdown (Schmitt and Sha, 2009; Lee and Schmitt, 2014). But to reiterate a point made in the Introduction, efficient word identification is the basic foundation that allows this kind of monitoring and self-correction to work.

Considering again the idea of evaluating both lower-level word processing and higher-level comprehension and expression, the proposed three-part series presents assessment options that include the range from:

- the *closed-ended* choice item (the cloze test),
- to *limited, or constructed, response* specific problem solving (the error correction test),¹
- to the traditional *open-ended* writing assessment with its unlimited number of satisfactory responses (the written expression task).

The limitations of each design might be able to be compensated for by results from one or both of the other assessment designs that present a different set of limitations, advantages and disadvantages. For example, trade-offs involve: different aspects of reliability, the transparency of task requirements (in the case of each assessment, what the target performance consists of), and the problem of making valid interpretations from results. See Fetsco and McClure (2005) and Miller et al. (2012) for discussion of the properties of different kinds of items and prompts. To re-emphasize, this idea is another proposal for further research that we still have not demonstrated in this study.

Returning to the question in the Introduction regarding the positive effect that evaluation methods should have on student learning, Amini and Ibrahim-González (2012) suggest from their findings that the cloze procedure, for example as a learning task integrated into the teaching program, compares favorably to other methods such as the post-reading comprehension question. Recall our recommendation on how to implement the cloze training activity in the section on procedures, in which students receive direct instruction on productive reading strategies. The error-detection/correction training activities (Appendix 6) are designed with this same purpose in mind. Studies of the effect of promoting awareness and monitoring at the sentence level have shown to improve higher level text comprehension (Wolter and Dilworth, 2013; Zipke et al., 2009).

APPLICATIONS FOR SECOND LANGUAGE LEARNING

Increasingly, students in Latin America are learning second languages in school, a minority or indigenous language (IL) as part of a language revitalization program, and/or English as a foreign language (FL), for learning a language of international communication.

Results of the present Ecuador pilot study continued work on a previous literacy assessment in the Mexico study (Chireac et al., 2019, in press) with an eye toward developing instruments for classroom use by teachers themselves. In the case of the multiple-choice cloze test, the new steam-lined design, with its ease of administration and scoring represent improvements over previous formats that, aside from practicality, might be more accessible for beginner and second language (L2) readers. The reason for this is that solution of items requires *recognition*, as opposed to *production* of a language form, retrieved often with more difficulty from long-term memory. For example, in the case of a given lexical item, the L2 learner may have mastery of some of the components (e.g., orthographic representation, an aspect of meaning) but others only partially or not at all (its phonological form, properties of syntax and morphology) (van de Craats, 2003). If the current multiple-choice format proves to be effective, parallel and equivalent assessments can be easily designed in the additional languages that form part of the L2 learning curriculum: FL-English and IL-Quichua, in the case of the highland region of Ecuador. The recently completed computer workshop facility, supported by the current project, is viewed by the community as an important resource for second language learning of Quichua, part of the above-mentioned language revitalization program.

Scoring the limited-response type error correction test is somewhat more demanding, requiring a judgment: the comparison between the student's response and the answer key's expected response. Similarly as on the cloze test, at least partial credit is obtained by recognizing the incorrect form. The writing assessment, of course, is based on a purely productive task.

Satisfactory results would be indicated by parallel progress across the grades when performance on the literacy assessments is compared between the languages. Work on the development of the parallel L2 assessment materials follows the recommendations in Ryan and Brunfaut (2016): where, often, the "ideal test writer profile is...spread across more than one person" (p. 394). Second language materials are parallel and equivalent, but not translations of the L1 version.

APPLICATIONS FOR MEASURING PROGRESS IN LITERACY

Chireac and Francis (2018) proposed an international study for evaluating the actual literacy achievement parity between boys and girls that should correspond to the region-wide Latin American gender parity in school enrollment at both the primary and secondary levels, historical attainment today confirmed by United Nations surveys of participation (attendance) in public school programs. Specifically, the proposal was for this evaluation to be carried out in rural community schools, with an emphasis on geographically distant and isolated localities for the purpose of confirming that actual parity in literacy performance corresponds to the reported enrollment parity, evenly, across all regional school systems. The report by Chireac and Francis, based on partial findings from three rural communities in Mexico and Ecuador, presented the working hypothesis that, in fact, such a gender parity has been attained, potentially measurable in actual performance on objective evaluation of literacy skills. Given that individual records of national school literacy testing, if they exist, cannot be relied upon in these communities, an on-site independent measure, taken by visiting evaluators, requires the use of an instrument of the type presented in this paper. For example, administered immediately following its corresponding training test, the multiple-choice cloze test provides for optimal conditions of consistency from one setting to the next. Parenthetically, the result of gender parity in performance on the cloze

assessment in the Results section (see Figure 1) is consistent with the results of Oliveira et al. (2011), mentioned earlier.³ Aside from applications of program evaluation, as in the previous two examples, classroom teachers should benefit from ongoing access to practical assessment tools tied to the learning objectives of their lesson plans.

NOTES

1. In the use of results from closed-ended type items (as in the cloze test of this study) and limited response items (the correction tests), different considerations present themselves. For example, for program evaluation purposes, comparing the overall performance between and among groups or studying the performance of the same class of students across time, scoring for the expected correct response (directly from the answer key) is reasonable because in the end the aggregate results and tendencies will not be appreciably affected. On the other hand, for providing students with feedback and for discussion of items, with the entire class or individually after the test, different conditions apply. For the closed-ended item, same as before, discussion of an incorrect response involves calling attention to how and why only one option is acceptable. In contrast, because in the limited-response design, more than one correct response is possible, evaluation of a non-expected response (one that is different from the answer key) will require closer assessment of each item. None of these constraints and conditions are defects, but rather (different) opportunities for teaching and reflection. All of the above applies, now in a different way, to scoring and providing students with feedback on the third, open-ended, assessment: The dear hunter [El cazador de venados] story.

2. “Parallel” in the second language is defined as progression across the grades, statistically significant from 2nd to 4th to 6th and onto the secondary grades. Significant (progressing) differences are always expected in the NL, language of daily instruction. Graphically, the L2 curves may not visibly appear as “parallel” to NL, as shown in Francis (2012: Figure 1), but they should be “parallel” statistically. Superior or exceptional results in L2 learning and L2 literacy would be indicated by converging progress (e.g., the gap in performance between NL and second language IL or FL, with time, actually remains constant or begins to narrow). A flat tendency across the grades (no significant differences) indicates questionable progress in L2 literacy and/or L2 learning overall. This result prompts educators to re-examine the bilingual or second language teaching approach. A flat and stagnant tendency across the grades suggests that performance in the FL or the IL does not even benefit from incidental transfer of knowledge and skills from literacy instruction in the NL. See Hamel (2016) for discussion of this relationship, specifically in regard to the IL educational context. Ideally, transfer of literacy-related language abilities should be evidenced “in both directions.” For example, progress in the first language (if the IL is the L1 of students) should favor academic language proficiency in the second language (the NL), and vice versa, effective instruction in the NL should favor preservation of the IL if the program is set up for this purpose. In this regard, the newly inaugurated San Lucas computer workshop presents the opportunity for digitizing the materials of Appendices 1–8 (together with future parallel IL versions), and reformatting them to create interactive learning modules.

3. Teachers and researchers are invited to administer the training activities and assessments in Appendices 1–10, and are welcome to publish the results, in addition to sending us a summary of the findings. Additional information regarding implementation of the training activities,

testing and grading procedures is available by writing to the authors. As mentioned in the Method section, none of the materials are appropriate for 1st grade testing. In addition, individual student scores cannot be used for either norm-referenced or criterion-referenced comparisons. As should be obvious, but to be absolutely clear, the score of a given student (e.g. in Bolivia) that falls above or below the average reported for grades 2, 4 and 6 in this article (results from three elementary grade classrooms in Loja Province, Ecuador) should not be used as an estimate of his or her individual literacy learning progress in school or for any other individual assessment purpose, formal or informal.

***Silvia-Maria Chireac** is Professor in the Department of Language Teaching and Literature at the University of Valencia. She is currently working on research that compares second language literacy learning in bilingual communities in Spain and in Latin America.*

Email: Silvia.Chireac@uv.es

***Norbert Francis** is Professor Emeritus of Bilingual/Multicultural Education at Northern Arizona University and currently Visiting Researcher in the Graduate Institute of Linguistics at Chengchi University. He works on problems of second language literacy and bilingualism.*

Email: Norbert.Francis@nau.edu

***John McClure** is Associate Professor of Educational Psychology at Northern Arizona University with research on the psychology of learning and motivation, statistics, and psychological assessment.*

Email: John.McClure@nau.edu

ACKNOWLEDGEMENTS

We thank professors Galo Rodrigo Guerrero-Jiménez, Norman Alberto González Tamayo, María de los Ángeles Guamán, Ángel Martínez de Lara, Carlos María Vacacela Medina and Bertha María Villalta Córdova of the Universidad Técnica Particular de Loja (UTPL) for their participation in the implementation of the pilot assessments in the community of San Lucas, and for their valuable observations and comments on earlier versions of the literacy assessments. We all express our appreciation to the UTPL for its support of the project: *Bilingüismo e interculturalidad, aprendizaje de segunda lengua y aprovechamiento escolar en la comunidad de San Lucas, provincia de Loja*, and to the Universitat de València for its support of the project: *Creación de laboratorio interactivo de lenguas vivas quichua-castellano*. Finally, we thank the Dirección General de Educación Indígena (México) for use of the “Cazador de venados” graphics, specifically to illustrator: Antonieta Castilla.

APPENDICES

Appendix 1

Cloze test

<http://jan.ucc.nau.edu/nf4/HistoriaBosque.pdf>

Appendix 2

Error correction test: Spelling

<http://jan.ucc.nau.edu/nf4/LaBrujaOrt.pdf>

Appendix 3

Error correction test: Grammar

<http://jan.ucc.nau.edu/nf4/LaBrujaGram.pdf>

Appendix 4

Error correction test: Punctuation/Capitalization

<http://jan.ucc.nau.edu/nf4/LaBrujaPunt.pdf>

Appendix 5

Cloze training activity

<http://jan.ucc.nau.edu/nf4/ReyMidasCloze.pdf>

Appendix 6

Error correction training activity

<http://jan.ucc.nau.edu/nf4/ReyMidasErrores.pdf>

Appendix 7

Model story (prompt) for writing assessment

<http://jan.ucc.nau.edu/nf4/CazadorDeVenados.pdf>

Appendix 8

Graphic context support for writing assessment

<http://jan.ucc.nau.edu/nf4/CazadorDeVenadosEscenas.pdf>

Appendix 9

Sample response paper (two-sided)

<http://jan.ucc.nau.edu/nf4/CazadorDeVenadosHoja.pdf>

Appendix 10

Writing Sample

<http://jan.ucc.nau.edu/nf4/CazadorDeVenadosWritingSample.pdf>

REFERENCES

- Amini, M. and Ibrahim-González, N. (2012). The washback effect of cloze and multiple-choice tests on vocabulary acquisition. *Language in India*, 12, 71–91.
- Bongartz, C. M. (2016). Bilingual and second language development and literacy: Emerging perspectives on an intimate relationship. *Selected Papers of the 21st International Symposium on Theoretical and Applied Linguistics (ISTAL 21)*, 3–42.
- Brown, J. D. (2013). My twenty-five years of cloze testing research: So what? *International Journal of Language Studies*, 7, 1–32.
- Chireac, S.-M., Francis, N. and McClure, J. (2019, in press). Awareness of language Literacy and second language learning of Spanish in Mexico. *International Journal of Bilingual Education and Bilingualism*
- Chireac, S.-M. and Francis, N. (2018). Alfabetización en la comunidad rural de América Latina: Las niñas en la escuela. *Contextos Educativos*, 21, 153–168.
- Dirección Nacional de Educación Intercultural Bilingüe (DNEIB) (2012). *Marco legal educativo: Constitución de la República, ley orgánica de educación intercultural y reglamento general*. Quito: Ministerio de Education.
- Ehri, L. (2014). Orthographic mapping in the acquisition of sight word reading, spelling memory and vocabulary learning. *Scientific Studies of Reading*, 18, 5–21.
- Ferris, D. and Hedgcock, J. (2005). *Teaching ESL composition: Purpose, practice and process*. Mahwah: Lawrence Erlbaum.
- Fetsco, T. and McClure, J. (2005). *Educational psychology: An integrated approach to classroom decisions*. New York: Pearson.
- Foorman, B. R.; Arndt, E. J. and Crawford, E. C. (2011). Important constructs in literacy learning across disciplines. *Topics in Language Disorders*, 31, 73–83.
- Francis, N. (2012). *Bilingual competence and bilingual proficiency in child development*. Cambridge: MIT Press.
- Fraser, C. (2004). Lire avec facilité en langue seconde. *La Revue Canadienne des Langues Vivantes*, 61: 135–160.

- Gombert, J.E. (2005). Apprentissage implicite et explicite de la lecture. *Rééducation Orthophonique*, 223, 177—187.
- Hamel, R.E. (2016). Bilingual education for indigenous peoples in Mexico. In O. García, A.M. Lin and S. May (Eds.), *Bilingual and multicultural education: Encyclopedia of language and education* (pp. 395—406), Bern: Springer.
- Lee, P. A. and Schmitt, M. C. (2014). Teacher language scaffolds the development of independent strategic reading activities and metacognitive awareness in emergent readers. *Reading Psychology*, 1, 32—57.
- Leikin, M.; D. Share and M. Schwartz. (2005). Difficulties in L2 Hebrew reading in Russian-speaking second graders. *Reading and Writing*, 18, 455—472.
- Lonigan, C. J.; Purpura, D. J.; Wilson, S.B.; Walker, P.M. and Clancy-Menchetti, J. (2013). Evaluating the components of an emergent literacy intervention for preschool children at risk for reading difficulties. *Journal of Experimental Child Psychology*, 114, 111—130.
- Miller, D.; Linn, R. and Gronlund, N. (2012). *Measurement and assessment in teaching*, 11th edition. London: Pearson.
- Mostow, J.; Huang, Y.-T.; Jang, H.-J.; Weinstein, A.; Valeri, J. and Gates, D. (2017). Developing, evaluating, and refining an automatic generator of diagnostic multiple choice cloze questions to assess children’s comprehension while reading. *Natural Language Engineering*, 23, 245—294.
- Oliveira, K.; dos Santos, A., Boruchovitch, E. and Marín Rueda, F. (2011). Cloze test in primary education: Differential operation of items. *Liberabit: Revista de Psicología*, 18, 45—52.
- Rogier, D. (2014). Assessment literacy: Building a base for better teaching and learning. *English Teaching Forum*, 3, 2—13.
- Ryan, E. and Brunfaut, T. (2016). When the test developer does not speak the target language: The use of language informants in the test development process. *Language Assessment Quarterly*, 13, 393—408.
- Samuels, S. J.; Ediger, K.; Willcutt J. and Palumbo, T. (2005). The role of automaticity in metacognition and literacy instruction.” In S. Israel, C. Block, K. Bauserman and K. Kinnucan-Welsch (Eds.), *Metacognition in Literacy Learning Theory, Assessment, Instruction, and Professional Development* (pp. 41—60). New York: Routledge.
- Schmitt, M. C. and Sha, S. (2009). The developmental nature of meta-cognition and the relationship between knowledge and control over time. *Journal of Research in Reading*, 32, 254—271.

- Share, D. and Stanovich, K. (1995). Cognitive processes in early reading development: Accommodating individual differences into a model of acquisition. *Issues in Education*, 1: 1—57.
- Shatil, E. and D. Share. (2003). Cognitive antecedents of early reading ability: A test of the modularity hypothesis. *Journal of Experimental Child Psychology*, 86, 1—31.
- Svalberg, A. (2007). Language awareness and language learning. *Language Teaching*, 40, 287—308.
- Tong, X.; S. Deacon and K. Cain. (2013). Morphological and syntactic awareness in poor comprehenders: Another piece of the puzzle. *Journal of Learning Disabilities*, 47, 22—33.
- Van de Craats, I. (2003). L1 features in L2 output. In R. van Hout, A. Hulk, F. Kuiken and R. Towell (Eds.), *The interface between syntax and the lexicon in second language acquisition* (pp. 69—95). Amsterdam: John Benjamins.
- Wolter, J.A. and V. Dilworth. (2013). The effects of a multilinguistic morphological awareness approach for improving language and literacy. *Journal of Learning Disabilities*, 47, 76—85.
- Yau, J.-L.C. (2009). Reading characteristics of Chinese-English adolescents: Knowledge and application of strategic reading. *Metacognition and Learning*, 4, 217—235.
- Zipke, M.; L. Ehri and H. Cairns. (2009). Using semantic ambiguity instruction to improve third grader's metalinguistic awareness and reading comprehension: An experimental study. *Reading Research Quarterly*, 44, 300—321.