Alternative Assessment: Opportunities and Challenges in Evaluating Reading

Guadalupe López Bonilla
Mara Rodríguez Linares

Translated by Trena Brown

Abstract:
This paper discusses the pertinence and applicability of alternative assessment in evaluating reading comprehension. In particular, it analyzes the limitations and scope of performance assessment, based on a study of high school students. A description is given of the design and application of an instrument intended to access the ways students approach the reading of two expository texts with historical contents. The instrument was applied to two complete groups of second-semester students in a public high school. The results allow reflecting on the need to explore these forms of assessment in contexts negotiated by the participants.

Key words: Reading assessment, alternative assessment, performance assessment, intertextuality, critical literacy.

Introduction
International reading assessments indicate that elementary school students in Mexico learn to decode a text without understanding its meaning; they also reveal that a high percentage of students who finish secondary school are able to carry out the most basic tasks in reading.
On addressing reading, the *Primer Estudio Internacional Comparativo en Lenguaje, Matemática y Factores Asociados para alumnos del Tercer y Cuarto Grado de la Educación Básica*, prepared by UNESCO, points out that in Mexico, most students in these grades are able to identify specific information from a text, although fewer students grasp the text’s meaning (UNESCO, 2000). In the case of the upper grades, the Program for International Student Assessment (PISA 2000)—sponsored by the Organization for Economic Cooperation and Development (OECD)—provides information on the reading ability of 15-year-old students in 32 countries, including Mexico. In contrast with the UNESCO study, which obtained results exclusively from multiple choice questions, the complexity of the instrument utilized in the Proyecto PISA—a combination of multiple choice and open-ended questions—permits distinguishing among six levels of reading competence. According to this study, only 1% of Mexican students attained the highest levels, while almost 80% of the sample from Mexico was able, in the best of cases, to perform basic reading tasks (30% of the total) or complete elementary tasks such as locating explicit information (OECD, 2001:45).

It should be mentioned that the PISA testing instruments for Spanish-speaking students made use of texts translated from other languages (primarily from English); therefore, in spite of the attention paid to their “cultural and linguistic equivalence” (Gil, 1998:9), they may not be culturally relevant for the Mexican sample due to their origin in non-Mexican contexts. The cultural relevance of texts employed to evaluate reading comprehension is fundamental for constructing instruments that are pertinent to specific contexts: as other writers have emphasized, all language processing activities require knowledge of the world, which is generally linked to particular situations (Alderson, 2000). "Knowledge may be limited since the worlds of other people function differently … therefore, cultural knowledge is crucial for understanding texts" (Alderson, 2000:45). We have no desire to minimize the results of the PISA project, yet we believe that cultural knowledge may be a problem that affects some of the questions on the OECD testing instrument. Although the instrument includes a high percentage of open-ended questions, its assessment criteria uses a binary system (0-1) that limits the possible interpretations of texts.

To document the reading competence of students in Mexico, it is necessary to carry out classroom research that takes into account the reading and writing experiences to which students are exposed, in addition
to the contexts and evaluation procedures that undoubtedly influence student learning.

In order to respond to some aspects of these questions, in August of 2000, we started work on a study with teachers and students from a public high school in the city of Tijuana, Baja California. On the most part, the school's enrollment of approximately four thousand is composed of middle-class youth from an urban context. Our research on site was facilitated by the willingness and interest shown by the school's administration and teaching staff.

During the study, we interviewed two literature teachers and two history teachers during the morning shift; we made two classroom visits, recorded on audio, in each teacher's classroom; we collected homework, exercises and examinations used by each teacher over the semester, and applied a survey to two groups of students for each teacher, in order to define sociocultural profiles and reading habits. The study's initial results permitted documenting the value that high school literature and history teachers attach to reading and writing as necessary experiences for academic success in the subjects taught. Given the orientation of the curriculum in the two subjects—a clear encyclopedic leaning in the history programs, and formal and structural aspects in the literature course—it is not surprising that teacher interviews, sample assessments and classroom visits verify that reading and writing as learning experiences do not occupy an outstanding place in teaching. The forms of evaluation are primarily multiple-choice examinations, which imply a substantial reduction of the curriculum, while reading occurs almost exclusively in the classroom; the appearance and correct handling of the most elementary linguistic conventions are the basis for evaluating the limited exercises that require students to express their reflections in writing. Even so, all the teachers agreed that their students' faulty reading comprehension skills and reading habits are a central problem in their experience as high school teachers (López Bonilla and Rodríguez, 2002).

On the other hand, studying the interviews within a conceptual and methodological framework of critical analysis of discourse (Gee, 1999; Fairclough, 1995), permitted identifying some of the teachers' theoretical assumptions regarding reading and their area of knowledge, as well as the cultural models and ideologies (Gee, 1999) that guide their teaching.

Given the need to document students' reading competence, another phase of the project consisted of the preparation, pilot study and application of an examination to evaluate the reading comprehension of high school students. The testing instrument utilized two expository texts with historical
content taken from books for high school students. This project presents the results of this phase of the study by addressing the exam characteristics and their pertinence in evaluating reading comprehension, as well as the process of preparing the testing instrument. Also described are the criteria utilized to evaluate the testing instrument, taking into consideration the complexity of evaluating instruments of this nature, in which reading habits neither appear in isolated form nor constitute discrete units. The discussion of results allows documenting the strategies students use, and identifies the aspects of questions they find most difficult.

**Reading in the School Setting**

Reading is a socially acquired practice (Gee, 1996); as such, it is governed by the cultural and institutional guidelines of a certain society. The scholastic institution fills a fundamental function in teaching individuals to read by orienting their forms of accessing, interpreting and building meaning when faced with texts of various types. Their reading depends on the purposes established in the scholastic context, and on the ways these practices are valued and evaluated by teachers. Students perceive the complexity of reading tasks according to certain rites of execution and orientation towards what is considered culturally “correct” (Greene and Ackerman, 1995).

The debate of recent decades has given rise to innovative proposals that view reading as a complex and dynamic process through which the individual constructs meaning by interacting with the written text (Langer, 1987; Pearson and Stephens, 1994). When seen in this manner, as an interactive process among the reader, the text and the context (Ruddell and Unrau, 1994; Rosenblatt, 1978), reading is inextricably linked to writing (Nystrand, 1986; Greene and Ackerman, 1995), and participants in the processes of comprehension include the text, the context, previous knowledge and the particular tasks of each situation in which reading occurs (Spiro, 1980; cited by Greene and Ackerman, 1995).

For this reason, the scholastic context of reading is no longer proposed as an isolated activity. It is understood that in order for the student to participate actively in the construction of knowledge, he must have access to multiple and varied opportunities to discuss and reflect on what he reads and writes (Heath, 1987)—activities that become meaningful in all subjects, rather than in only language classes or reading and writing classes (Pearson and Stephens, 1994). This idea has been translated into
curricular modifications leading to constant efforts by teachers and the students themselves to build particular reading and writing models and reflect critically on the implications for educating literate individuals (Weade and Green, 1989).

In the evaluation of reading and writing, educational reforms in the international setting reflect a tendency to utilize methods other than standardized multiple-choice examinations, which have proven to be poor indicators of academic performance in general and of reading in particular (Popham, 1999; Linn and Herman, 1997; Gee, 1992); using multiple-choice examinations as the sole means of evaluation entails a mechanical view of learning (Shepard, 2000). As a result, there has been increasingly frequent use of portfolios, of "authentic" evaluation in the classroom and of examinations with open-ended questions through which the student shows in writing his skills, previous knowledge and experiences with reading comprehension (Harrison and Salinger, 1998).

In the case of Mexico, the curricular reform of 1993 for teaching language in elementary education integrates reading and writing under the heading of written language; however, in the teaching materials of 1996, they appear as separate activities (Carrasco, 2000). On the other hand, although this curricular reform attempts to center learning on the subject, the proposal emphasizes that "carelessness in forming readers continues to be evident" (Carrasco, 2000:99).

In secondary education, the curricular reform of 1993 involves contradictions between its discourse and modifications, with regard to the content and forms of viable evaluation under current teaching conditions (Quiroz, 1999). As an example, Quiroz points to the incompatibility between the purposes of history programs (the identification of the principal characteristics of eras, the notion of historical process) and program contents, with an encyclopedic vision of history that prevents the proposed analysis and reflection (Quiroz, 1999). According to Quiroz' research, something similar happens in the programs of study for Spanish: minimal modifications in terms of content, in spite of the explicit emphasis on applying knowledge in oral and written communication. If, in addition, the teaching conditions of most secondary teachers are taken into account—a high number of class hours with large groups—it becomes clear that the purposes of the reform are far from becoming a reality, and that teachers turn to "routines of standardized tests, since they are the only ones possible under current working conditions" (Quiroz, 1999:115).

At the high school level, the diversity of study programs impedes a detailed analysis of the curricular proposal. However, it is worthwhile to mention
that the reading and composition courses of the first and second semesters center on a prescriptive and fragmentary focus on reading and writing, and literature courses emphasize the formal aspects of literary genres without necessarily reading, discussing and analyzing works typical of each type.

Given this panorama, it is not surprising that recent studies on the academic quality of high school students indicate that their reading comprehension levels are similar to those of elementary school students (Benítez, 1999)—a situation that may influence the causes of failure and drop-out rates (Tapia, Tamez y Tovar, 1994), as well as poor academic performance in the upper grades. In particular, a study carried out by Díaz Barriga on critical thinking skills among high school students emphasizes “their poor comprehension of historical texts” and their limited experience in answering open-ended questions (Díaz Barriga, 2001:548).

Multiple-choice examinations are favored in evaluating learning, and their indiscriminate use for promotions and entering the upper grades does not contribute to educating literate individuals. The study of alternative forms for evaluating reading comprehension is necessary in order to build contexts for learning in which students can become aware of their participation as active readers (Solé, 2001:13).

**Alternative Practices for Evaluating Reading**

If reading is conceived as an interactive process through which multiple and flexible meanings are constructed, then evaluating reading—inaccessible due to its very nature—becomes impossible. As Pearson, DeStefano and García (1998) affirm, in the past, the assessment of reading has consisted of developing the best possible forms to make inferences about what happens during the act of reading (Pearson et al., 1998). Some point to the incongruity of this evaluation as a project with modern logic put into practice in a postmodern society (Harrison, Bailey and Dewar, 1998). In the end, evaluating reading comprehension favors a particular form of reading that is understood to be “correct” or “adequate”. In other words, reading practices are accompanied by interpretations that are not always explicit, but always political (Gee, 1996). The point in question is the meaning of reading comprehension, a construct that is increasingly ungraspable due to its fluid nature and location in specific contexts (Harrison et al., 1998; Pearson et al., 1998).

It is in this context that the forms of alternative assessment have become popular in the past decade. In contrast to traditional practices, in which the
student selects a response from a menu of possibilities (multiple choice, true or false, matching), alternative practices require the student to construct a response, product or demonstration. According to Winograd, Blum and Noll, alternative assessment is understood to occur daily in significant learning contexts. Such practices reflect experiences conceived to be "authentic" and documented, among other methods, through observations, journal entries, portfolios, experiments and performance assessments. Emphasis is centered on the self-reflection of the subject and his comprehension of what is done without considering the recall of isolated data (Winograd et al., 1999).

Some authors have pointed out that these methods fill at least two functions: they stimulate discussion between teachers and students while marking milestones in student learning by providing a torrent of descriptive evidence of student comprehension and progress (Winograd et al., 1999). Even so, such forms of evaluation do not escape from the same criticism that has cast doubt on the validity and credibility of traditional examinations and tests. For example, in the case of open-ended questions, there are inevitable problems of subjectivity and interpretation from at least three sources--the person who prepares the question, the person who answers the question, and the person who evaluates the answer--while tasks may not always be clear for the person who must carry them out (Pearson et al., 1998). The challenge consists of retaining reader subjectivity as something positive and valuable, without going to the extreme of absolute relativism, which detracts from the nature of the texts themselves. For example, if both the questions and criteria of evaluation are prepared and discussed beforehand in the classroom context, and if the evaluation is carried out by the students, the experience becomes an opportunity for students to reflect on reading and to adopt a critical position.

A clear example of the problems encountered by performance assessments in evaluating reading is provided by Goldberg and Kapinus' research on the reading assessment program in the state of Maryland. By analyzing the unexpected responses obtained from a random student sample, Goldberg and Kapinus identified various typical errors, especially errors of context (discrepancy between the context of the evaluation and of the classroom), errors of the testing instrument (lack of clarity in the instrument's instructions or language), and errors of the person answering the questions (incomplete or inadequate responses), as well as the combination of two or more of these errors (Goldberg and Kapinus, 1993).

In spite of the limitations of this type of instruments and of alternative assessment in general, all these authors agree that these forms of
evaluation offer opportunities for reflecting on the act of reading, regardless of the purposes and contexts of evaluation. Thus, a performance assessment with educational purposes carried out in the classroom will lead to results that are very different from those obtained from a similar instrument given for scholastic promotion, entry to upper grades or rendering of accounts (high-risk examinations).

Although replacing traditional methods of evaluation with alternative methods does not in itself guarantee better teaching (Valencia and Wilson, 2000), it is true that changes in teaching cannot be encouraged if the current forms and contexts of evaluation are not modified (Miras, Solé and Castells, 2000).

Because of the advantages offered by alternative assessment and in particular by performance assessments, we found it appropriate to prepare an instrument of this nature to comply with the purposes of our research: to access the ways that high school students approach the reading of academic texts. Aware of the limitations implied by an exercise of this type, we decided to construct an instrument with two expository texts on the same topic, but with different postures. We consider it important to explore this exercise since a reader faced by multiple texts on the same topic can make intertextual connections and acquire a critical posture with regard to what he has read (Beach, Appleman and Dorsey, 1994). Our intention was to discover the positioning of students with regard to possible intertextual reading, and to determine if they questioned the text's authority.

The Testing Instrument: Preparation Procedure

Performance assessments must have at least three characteristics: 1) Students construct responses instead of selecting them. 2) Evaluation formats allow teachers to observe student performance based on specific tasks. 3) Evaluation criteria reveal the students' patterns of processing and learning (Fuchs, 1995).

In constructing our instrument, we considered potentially significant texts of cultural relevance for the participants. The criteria for selecting the reading material were as follows:

- Authenticity: Texts were chosen that could be transcribed integrally, and that had not been written expressly for evaluation purposes.
- Level: The texts were taken from history books used by high school teachers and students. The two explanatory texts chosen were
extracted from books used at other schools. Both are from books published in Spanish: the first is about United States history, written by United States authors, while the second is about Mexican history and its author is Mexican. The sources were specified at the bottom of each reading. Neither of the readings is utilized by the teachers who participated in the study.

- Topic: Two versions of Texas' war of independence were selected, in order to induce the student to contrast the versions, to distinguish and compare points of view, to differentiate between facts and authors' interpretations, to weigh the validity of arguments and to identify the author's posture.

- Length: The one and one-half pages of the first reading and the single page of the second can both be read by students in ten to fifteen minutes (Annex 1).

**Criteria for Preparing Questions**

The open-ended questions were prepared under a criteria of "authenticity". Open-ended questions have been defined as those that have no preconceived response; they accept a wide variety of responses and allude to concepts, processes and skills that surpass the declaratory content of the subject in question (Badger and Thomas, 1992).

This type of questions requires greater efforts by teachers and evaluators when interpreting the responses, and implies the adoption of alternative methods of assessment that go beyond the simple classification of responses as “acceptable or unacceptable”, and/or “correct or incorrect”. With these criteria, three questions were prepared: two short-answer questions and one long-answer question.

Once the questions had been formulated, a pilot study was carried out with three groups of high school students from other schools (middle-class urban high schools), in which a total of 123 third-semester high school students participated. Particular attention was paid to the time students took to read the texts and write their answers, to the clarity of the task, and to the level of difficulty of each question. The final version of the questions is as follows:
Short-answer question 1: “In what form does reading these two passages allow you to have a better understanding of Texas' situation before declaring its independence? Comment.”

Short-answer question 2: “Comment on the form that each passage describes Antonio López de Santa Anna. Are there differences? What does this tell you about the posture of the author of each passage? Write a commentary.”

Long-answer question: “Compare the form each author narrates the battle of the Alamo. Describe the coincidences and differences in the two passages. Reflect on the two versions and explain the reason for these differences.”

The testing instrument was applied to two complete groups of second-semester high school students, all regular students in the morning shift (95 in total). The characteristics of the students who participated in the study are shown in Chart 1:

<table>
<thead>
<tr>
<th>Chart 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student Characteristics</td>
</tr>
</tbody>
</table>

*In the columns that indicate the parents' education, each number corresponds to the highest level of studies of the father and the mother. Level 1 corresponds to elementary school, 2 to middle school, 3 to high school, 4 to technical studies and 5 to university studies.*

The instructions were to read the selected passages, answer the two multiple-choice questions at the end of each text, and respond to the three open-ended questions (two short-answer and one long-answer) that involved both passages. The examination was given in the classroom during the hour scheduled for history.

Criteria of Evaluation

The technique most utilized for evaluating performance assessments consists of adapting descriptive scales that contain specific criteria for classifying responses at each grading level, with a description of the responses that could pertain to each grading level (Moskal, 2000).

Two general methods of evaluation can be distinguished: the analytical method, which uses a scale to evaluate each aspect involved in carrying out a task (for example, a separate evaluation of the ability to integrate previous knowledge, or the quality of argumentation); and the holistic method, which uses criteria considered jointly in a single descriptive scale (when separating the evaluation into independent factors is impossible).
Models with a holistic scale serve as support in making broader judgments concerning the quality of the process or the product of academic effort (Moskal, 2000).

Both models, however, have reported inconsistencies among readers in types of evaluation based on unclear and/or ambiguous criteria. Thus readers may find it difficult to distinguish between “an adequate understanding of the text” and “a complex answer that shows profound understanding of the text” (Resnick, Resnick and DeStefano, 1993).

There is controversy regarding the possibility of identifying and labeling isolated comprehension skills. Reading skills evidently do not form discrete units, but always appear as "superimposed". On the other hand, it is impossible to verify the existence of such skills directly. Difficulties of this type may be surmounted through a very precise description of the constructs— he skills hoped to detecta– s well as a detailed explanation of the criteria adopted to put these theories into operation (Alderson, 2000).

In light of these considerations, this study adopts holistic criteria with a scale of five grading levels to evaluate the short-answer questions, and six levels to evaluate the long answer. Each level reflects the degree to which the response complies with the question's purposes, and the level describes the skills that correspond to that grade. When we designed the questions, we made note on the scale of the skills we hoped to detect, and on finishing the pilot study, we reviewed the constructs and made the pertinent corrections. In this manner, the test allowed us to establish a relationship between the different levels of the scale and the skills manifest in each level; it also served to take samples of student responses, to describe their characteristics and assign them a grade.

It is important to point out that the scale does not emphasize a particular interpretation of the passages, but notices what students do or fail to do when faced by the tasks required by each question. As a sample, Annex 2 reproduces the final version of the scale for the short-answer questions 1 and 2.

In the long-answer question, the levels of the United States Educational Evaluation Service are used to evaluate reading comprehension with responses of this nature (Langer et al., 1995).

All the responses were evaluated by three readers. The Pearson range of correlation among readers was 0.73 to 0.91. In all the cases, disagreement was ± 1, with a high percentage of total agreement among readers.
Results

As shown in the questions and in the skills necessary for answering them, the questions required the students to relate and connect ideas inside and outside of each passage in order to be able to identify different postures on the same historical event; they were also required to reflect on each posture and adopt a critical posture with regard to the reading. However, on analyzing the responses, we were able to prove that a high percentage of students was not able to reflect on the tasks necessary for answering the questions, nor think about the possible application of the acquired knowledge. Such is the case of the responses to short-answer question 1 (“In what form does reading these two passages allow you to have a better understanding of Texas' situation before declaring its independence?”), which resulted in a significant number of "inadequate" responses, as shown in Graph 1.

Graph 1
Short Answer Results 1

For example, some typical responses follow:⁵
- "Narrated in language that is understood and that has colloquial language”
- "It has a good structure, good coherence, words that we can understand. It has a title, a subtitle, background, development and a conclusion.”
- "Because of the structure the text shows in coherent form.”

In these examples, the students do not distinguish between one passage and the other, and their comments point to formal aspects. The responses
make evident that their scholastic experience favors, in the best of cases, the study of the formal characteristics of texts without paying attention to the meaning and the construction of meanings in the process of reading. Therefore, it is not surprising that students were unable to interpret the meaning of the question, not even literally, and commented instead on the formal aspects of the passage—a confirmation that in the school setting, students interpret tasks and activities according to the "rites of execution" and cultural orientation perceived as adequate (Brooke, 1987; Nelson, 1990; cited by Greene and Ackerman, 1995). In this sense, we can state that these responses are examples of “errors of context” (Goldberg and Kapinus, 1993), since it is clear that the demands of the question do not correspond to student experience.

On the other hand, a minimal percentage of students could formulate responses that evidenced their performance of the tasks required by the question: to distinguish between versions of the same topic, to reflect on the elements provided by each version, and to think about the possible application of the acquired information. Examples of responses in compliance with these requirements follow:

- "They allow me to have a better understanding because they both talk about the same topic from different focuses and/or perspectives. One author gave it more importance and lengthened some topics or events that he considers more important, and the other author did the same. Reading both passages offers us a wider and clearer panorama of what happened."

- "It helps you know more because it gives you two different focuses of what really happened, and in this way we have the opportunity to understand the reasons or motives for the independence of Texas, by having two opinions and making our own."

- "Because it gives us the two points of view of the winners and of the losers."

As these three responses show, with their varying ideas and shades of meaning, the readers assumed a critical posture that questioned the text's authority.

The responses to the question, “Comment on the form that each passage describes Antonio López de Santa Anna. Are there differences? What does this tell you about the posture of the author of each passage?” were inadequate for 28%. In these cases, the students were unable to identify information expressed explicitly or to integrate information from different texts. The following are examples of this type of responses:
- 'There are no differences. The authors may describe it in different ways but they will always say the same, since Santa Anna only thought about himself and his wealth."

- 'In the text done by Brom Juan it says that Santa Anna was defeated and [taken] prisoner. They called him Napoleon of the West and he was very ambitious."

As the first response shows, the student ignores the content of the passages and starts from his previous knowledge—his idea of Santa Anna—to assume that all authors share his view of this historical figure. It could even be stated that for this student, reading the passages is not a necessary activity since he assumes their meaning based on his previous knowledge and on what he perceives as correct. His responses make manifest an elaborated inference that deviates from the text—a practice that tends to occur when there is disagreement between the reader's scheme of knowledge and the text itself. In these cases, the reader's desire to match the text with his own knowledge may lead him to incur in a "distortion" of the text (van Dijk and Kintsch, 1983:52). It is of interest to underline that this response is similar to the conception of historical texts expressed by one of the history teachers in the interview: it does not matter which one is used because they all say the same thing.

On the other end of scale, we obtained only one response that provides evidence of performing the question's tasks:

- 'The first one describes Santa Anna's personality more and according to the author, describes him with a feeling of rancor. It says in general terms that he was a bad person.

  In contrast with the first, the second one does not talk about Santa Anna's personality, and does not give it importance. I think that the above is because the first author is from the United States and the second is from Mexico."

In this case, the student addresses the passages separately and points to traits that he considers important, attributing them with a meaning—a feeling of rancor”, “does not give it importance” and a cause: “the first author is from the United States and the second is from Mexico."

The total distribution of the responses is illustrated in Graph 2.
The responses to the long-answer question were similar to those of the second short-answer question: 66% of the responses were below level 3, and 25% of the answers were inadequate, as shown in Graph 3.

The following are examples of inadequate responses:
- **TEXT 1**: The first passage narrates more in depth how a certain number of United States people who emigrated to Texas became associated, and how they expanded in the territory, it also talks about their relationships with the inhabitants of the same, in this passage all the events that brought about the Independence of Texas are explained little by little and in detail.

The second passage talks directly about the division of the Mexican territory, and more dates and names that refer to this event are mentioned, as well as more about the president of that time, Santa Anna, and Mexico's opinion about the independence of Texas.

- **TEXT 2**: The first passage gives more specific data about the lifestyle of Mexicans established in Texas, and secondly, it tells us how the United States wanted to change the Indians who lived in Texas to change them, since they were convinced that their way of administration and trade was the best and the most useful.

The first passage also explains to us that Santa Anna was ambitious, which makes us think that he was not at all interested in the progress of Mexico, although he did fight for Mexico by killing 188 Americans to save Texas and caused more conflict. [Not] until Santa Anna was defeated and caught was Texas considered a sovereign state of the USA.

Both responses fail to show clear evidence of an initial understanding of the passages, and even ignore the question by not addressing directly the battle of the Alamo. In other words, there is no evidence that the students can locate specific information or much less prepare a commentary about it. Once again, these responses may be due to errors of context.

**Conclusions**

The results obtained from applying the testing instrument document students' limited experience in locating specific information, in relating ideas inside and outside of the text, and in identifying different points of view of a single topic. However, it is necessary to point out that an important cause of the high percentage of “inadequate” responses is attributable to the “errors of context” discussed above. And in spite of the efforts to utilize a testing instrument in agreement with the principles of alternative assessment, it is obvious that this exercise was not an authentic evaluation: it arose from a context foreign to the institution in which it was carried out.

These partial results, limited to a particular population and context, reveal that most of the students lack skills that would allow them to make
connections from diverse sources and adopt a critical posture with regard to their reading.

The reasons may be numerous, but we believe that the orientation of high school curriculums as well as of institutionalized evaluation must be seriously questioned. In this senses, the reflections of D. Hartman are of use in pointing out that an evaluation centered on the reading of a single text assumes that comprehension “is the act of reading isolated texts”, without taking into account that when good readers read, they connect ideas with their previous reading experiences to establish a network of associations that cause comprehension to go beyond the isolated text (Hartman, 1994:616).

Because of the complexity of evaluating reading, the exercise discussed in this paper undeniably has limitations. Yet it also has the advantage of disclosing the criteria for the evaluation and thus making explicit the type of reading favored. In our case, we have hoped to stimulate an intertextual posture in the reader and question the authority of the text.

The intention of this research has been to contribute to the discussion of new forms of evaluation that allow both teachers and students, in the classroom, to orient the practices of reading and writing. We believe that this type of instrument, more than a student evaluation, is an exercise that enables reflecting on the varied and enriching uses of written language. If teachers as well as students establish clear, common criteria for the forms of accessing, understanding and interpreting each text, and if inside and outside of the classroom they experience reading as a continuous dialogue about texts, the contexts from which texts arise and the ways texts acquire meaning, then reading and writing will no longer be seen as isolated activities that are meaningless for the young people of our country. For this reason, we believe that the classroom use of instruments similar to the one proposed here, as well as other methods of alternative assessment employed in a context of negotiation between teachers and students, will contribute to the formation of critical readers.

The possible success of exercises of this type does not depend on merely substituting one form of evaluation for another. On the contrary, studies that address the effects of alternative assessment in the classroom emphasize that in order for the evaluation to affect the learning process in a positive manner, it is necessary to train teachers, review the curriculum carefully, and analyze the policies that influence the predominant educational standards and forms of evaluation (Shepard, 1989; Pearson et al., 1998; Valencia and Wilson, 2000; Hiebert and Davinroy, 1993; Borko et al., 1993). In Mexico, it is necessary to keep these recommendations in
mind within the current context, in which national educational policy emphasizes evaluation and the rendering of accounts.

Notes

1. The UNESCO laboratory instrument included an open-ended question to evaluate written production, but it was not included in the results (UNESCO, 2001:182).

2. The PISA instrument located each reactive on a scale from 1 to 5, where 1 refers to the ability to make simple connections in familiar texts, and 5 to the ability to adopt a critical posture—readers who are able to formulate hypotheses, integrate information from diverse sources and demonstrate in-depth comprehension of a text. The sixth level is for students who are below level 1: in the case of Mexico, 16% of the sample (OECD, 2001:45).

3. This would be the case of the text utilized in reading unit 2 of the pilot study, questions 5 and 6 (OECD, 2000:27-29).

4. The term in English, literacy, translated into Spanish in some cases as “alfabetización,” and in other cases as “formación lectora” (OCED, 2000), is the closest to this posture. In the absence of an equivalent term in Spanish, we opted to refer to reading and writing as intrinsically related processes.

5. All the samples in Spanish are literal transcriptions of responses.

Annex 1

TEXT 1

I WESTWARD EXPANSION AND ECONOMIC GROWTH (1824-1848)

Colonization of Texas. In the 1820s and 1830s, a number of US residents—mainly southerners—took Mexico’s liberal colonization law to the letter and immigrated to Texas. With the help of slaves and cotton gins, they cultivated the fertile land and managed their businesses under the protection of Stephen Austin and other entrepreneurs who had contracted with the Mexican government to establish a certain number of families in Texas, in exchange for land concessions.

Over three centuries, the Spanish government had taken no more than 4,000 persons to Texas. Under the new conditions, the population of only Austin’s communities expanded from 2,000 in 1828 to more than 5,500 three years later. By 1836, more than 25,000 white men, women and children were scattered between the Sabine River and San Antonio de Bejar. The colonists
from the United States greatly outnumbered those descended from the Spaniards.

Friction between the Mexicans and Americans in Texas was probably inevitable. The Mexicans, long accustomed to Spanish proceedings, were naturally unprepared for the types of administrative and legislative responsibilities so easily assumed by the immigrants. The United States citizens living in Texas, aggressive and self-confident, were sure that their way of life was the most free, healthy and happy and in all senses superior to that of the Mexicans. They considered themselves, individually and collectively, as the adequate agents to impose reforms and advances on what they judged to be an uncivilized society burdened by generations of superstition and indolence. The newcomers were generally unable to recognize the spirituality and manners of the Spanish culture and criticized the Mexican peasants as illiterate and unaware of the outside world. They also ignored the equally pertinent truth that the Mexican peasants as well as their masters were sensitive and proud.

The Americans in Texas were annoyed by the changes in Mexican policy and by the uncertainty of their own situation. The Mexican government seemed indifferent to educational needs and to the application of the law, and did nothing to encourage the Americans' hopes that Texas could separate from the state of Coahuila to which it had pertained for so long. This negligence, as well as divergent attitudes about religion and slavery, contributed to widening the existing gap between native Mexicans and the immigrants who had come from the north.

The Fight for Independence. Meanwhile, in Mexico City, a growing tendency towards a dictatorial government was reducing the chances for conciliation. The incarnation of this despotic spirit was Antonio López de Santa Anna, who became the president of Mexico in 1833. Santa Anna, the Napoleon of the West, was ambitious, inclined to intrigue and also an efficient commander, as long as luck favored him. As president, he inflexibly crushed any indication of liberalism in Mexico's central government, and then focused his attention on Texas, where the Americans were protesting vehemently against this president's abandonment of the "enlightened" Mexican Constitution promulgated eleven years earlier. The Texans' protests culminated in a proclamation of independence from Mexico on March 2, 1836.

Four days later, Santa Anna and the Mexican forces entered Texas and killed 188 Americans at the Alamo mission in San Antonio. Davy Crockett, Jim Bowie and William B. Travis were among the United States heroes who died defending the Alamo in San Antonio. That same month, in Goliad, on the south bank of the San Antonio River, James Walker Fannin, gravely wounded, surrendered the small troop he commanded to the Mexican general, José Urrea, believing that the Texans would be subject to the humane treatment normally granted to prisoners of war. Instead, Urrea, following orders from Santa Anna, executed most of the prisoners in cold blood, and then Colonel Fannin. If the
The cries of "Remember the Alamo!" and "Remember Goliad!" served for a long time as shouts of unification for the Texans.

During Texas' fight for independence, the forces of the young republic were under the skillful command of General Sam Houston. This robust son of Virginia had moved at a young age to Tennessee, fought under the command of Andrew Jackson in the War of 1812, studied law, and served as congressman and governor of Tennessee. He became established in Texas after Jackson sent him there to negotiate a treaty with the Indians in 1832.

Houston showed his leadership abilities in the Battle of San Jacinto, which was fought on April 21, 1836, less than two months after the defeat of the Texans at the Alamo and Goliad. Houston's troops surprised and defeated those of Santa Anna and took the Mexican dictator himself prisoner. Houston forced Santa Anna to sign the Treaty of Velasco, under which Mexico agreed to remove its troops from Texas and recognized the Río Grande River as the southwest border of the new Republic of Texas.

Taken from: Holman, Hamilton, "La democracia y el destino manifiesto" in Degler, Cochran y otros, Historia de los Estados Unidos. La experiencia democrática. Editorial Limusa, México, 1991. Pp. 204-205

TEXT 2

II GOVERNMENTS, INTERNAL AND INTERNATIONAL CONFLICTS (1821-1855)

Wars and Uprisings
Mexico is reduced by half. The English colonies of North America had become the United States Republic four decades before the independence of the Spanish possessions. The new country soon began expanding its territory, based on manufacturing and vigorous trade, and fed by strong immigration from Europe.

Many United States citizens were sincerely convinced of having attained the best form of society and government, and wanted to include all peoples, especially on the American continent. In practice, their attitude served as "justification" to take control of new territories, expel or annihilate the indigenous population, and for government officials to accumulate great earnings by speculating with the land they appropriated and sold to the new colonists, many of whom used slave labor.
The United States took advantage of the wars among European nations to acquire several of Europe's former colonies. In 1893, the United States bought Louisiana—an extensive territory to the west of the Mississippi River that included the important port of New Orleans—from Napoleon's government. The United States later occupied the Spanish colony of Florida (1812), claiming that Spain had allied with England, which the United States was fighting against that year; seven years later, the United States legalized the ownership of the territory by purchasing it from Spain. The next objective was northern Mexico, from Texas to the west coast, in order to extend its territory from the Atlantic to the Pacific.

**Texas.** The Hispanic population in Texas was sparse, and the American colonization of Texas had started in the colonial period. The Spanish government, as well as the Mexican government in later years, promoted this immigration, with the condition that immigrants were to be Latin and Catholic. The government of Mexico tolerated the violation of the ruling, and also authorized the colonists' use of slaves. United States and Mexican officials amassed fortunes by speculating on land concessions.

On various occasions, the Mexican authorities attempted to slow the arrival of more Americans and encourage Mexicans, but they failed in their efforts. The inhabitants of Texas requested statehood, in order to enjoy a degree of autonomy, but were refused by Mexico's president, Santa Anna.

Contradictions became more severe and culminated in Texas' declaration of independence. The United States government encouraged the independence movement by supplying the insurgents with weapons and resources. The Mexican army, headed by Santa Anna, fought and won some victories over the rebels, most importantly at the Alamo. But in the end the Mexicans were defeated and Santa Anna taken prisoner. Mexico withdrew its troops (1836), without accepting Texas' status as an independent nation. One year later, the United States recognized Texas as a sovereign state, soon to be followed by France and England, which promoted the strengthening of a new country between Mexico and the United States to hinder the expansion of what they considered the future threat of American power.

Taken from: **Brom, Juan, Esbozo de historia de México,** Editorial Grijalbo, México, 1998. Pp.171-172
Annex 2

Scale for short-answer question 1:
Level 4: The response covers satisfactorily all of the question's aspects.
  ▪ It distinguishes between different versions of a single fact.
  ▪ It reflects on the elements provided by each version.
  ▪ It thinks about the possible application of the acquired information.
Level 3: The response covers satisfactorily most of the question's aspects.
  ▪ It distinguishes between different versions of a single fact.
  ▪ It thinks about the possible application of the acquired information.
Level 2: The response covers satisfactorily some of the question's aspects.
  ▪ It distinguishes between different versions of a single fact.
Level 1: The response is inadequate.
Level 0: It does not answer the question.

Scale for short-answer question 2:
Level 4: The response covers satisfactorily all of the question's aspects.
  ▪ It identifies information expressed explicitly. It
  ▪ Integrates information from different texts.
  ▪ It identifies the ideas the author is trying to emphasize.
  ▪ It formulates hypotheses on topics, concepts, underlying evidence.
Level 3: The response covers satisfactorily most of the question's aspects.
  ▪ It identifies information expressed explicitly. It
  ▪ Integrates information from different texts.
  ▪ It identifies the ideas the author is trying to emphasize.
Level 2: The response covers satisfactorily some of the question's aspects.
  ▪ It identifies information expressed explicitly. It
  ▪ Integrates information from different texts.
Level 1: The response is inadequate.
Level 0: It does not answer the question.

Bibliography


OCDE (2000). *La medida de los conocimientos y destrezas de los alumnos. La evaluación de la lectura, las matemáticas y las ciencias en proyecto PISA 2000*.


*Article received:* June 19, 2002  
*Accepted:* February 7, 2003