This sampler of student responses to mathematics and science items was assembled to show how student responses differ depending on whether students are tested in English or in their native languages. The sampler aims to promote awareness among teachers and assessment developers that low language proficiency in English is a factor that can penalize the scores of English learners in science and mathematics assessment. The document consists of 5 parts. Part 1 describes the approach used to develop the sampler. Part 2 presents a summary discussion of the quantitative and qualitative analyses showing the degree of concurrence in teacher ratings of students responses within the same language and across languages. Part 3 discusses the implications of the results and offers recommendations, and part 4 presents suggested exercises designed to promote the use of the sampler as a professional development tool. Part 5 shows the student work samples. A total of 124 students and 8 teachers participated in this project. Four open-ended science items and 3 open-ended mathematics items from the National Assessment of Educational Progress were selected, and 3 native languages, Spanish, Chinese, and Haitian-Creole, are represented. Five appendixes discuss the scales used, scoring rubrics, data analyses, and sampler development. (Contains 19 references.) (SLD)
Testing English Language Learners: A Sampler of Student Responses to Science and Mathematics Test Items
Acknowledgements

This report is the product of the Council's State Collaborative on Assessment and Student Standards (SCASS) LEP consortium in collaboration with the Project to Improve Achievement in High Poverty Schools of the Resource Center on Educational Equity. Julia Lara and John Olson coordinate the Consortium activities. Support for this project comes from the members of the SCASS states and the United States Department of Education (USDE).

The LEP Sampler is the outcome of a process during which we engaged several teachers and LEP student education experts in a number of activities. Julia Lara wrote and edited sections of the Sampler and coordinated the activities that resulted in the development of this document. Guillermo Solano-Flores, WestEd, helped to conceptualize the effort, formalized the design of the study, led the item scoring session of the project, and was the lead writer of this report. Ursula Sexton, WestED, assisted during the scoring training session and in the substantive editing of the work sample commentaries. Cecilia Navarrete provided valuable insight for the initial conceptualization of the project, conducted the student work-sample review session during which the commentaries were written by the teacher-consultants, and wrote Part IV, "Suggestions for using work samples." Finally, Min Li, Stanford University, conducted the statistical analyses.

The Council staff acknowledges the work and support of the teacher-consultants whose names are listed in Appendix E of this document. These teachers exhibited great professionalism and commitment to advancing the education of LEP students.

Finally, financial support for the development of this document would not have been secured without the leadership efforts exerted by Sharon Saez, during her tenure at the USDE.

The preparation of this document was financed partly by funds provided by the U.S. Department of Education (U.S. Department of Education, under Contract #EA94052001). No official endorsement by the U.S. Department of Education or the Council of Chief State School Officers is intended or should be inferred. This report is intended to provide suggestions that school systems may find helpful in addressing the issue of inclusion of LEP students in large-scale assessments.

Copyright © 2001 by the Council of Chief State School Officers, Washington, DC.
# TABLE OF CONTENTS

Acknowledgements ........................................ ii

The Council's Assessment Center .......................... 1
The Council's Resource Center ............................ 1
The SCASS Program ......................................... 1
The SCASS LEP Consortium ................................. 2
Introduction .................................................. 3

Part I
The Sampler Project ........................................ 9
   Levels of Analysis ...................................... 9
   Native Languages ........................................ 9
   Participants ............................................ 10
   Item Selection ......................................... 10
   Item Translation ....................................... 10
   Test Administration .................................... 12
   Scoring .................................................. 12
   Data Collection ........................................ 14
   Design of the Sampler .................................. 16

Part II
Summary of Results ........................................ 19
   Quantitative Analysis .................................. 19
   Qualitative Analysis .................................... 20

Part III
Implications and Recommendations ....................... 21

Part IV
Suggestions for Using the Work Samplers ................. 24

Part V
The Sampler ................................................. 30

Appendices
A. Suggested Fluency and Literacy Scale ................ 179
B. Scoring Rubrics ........................................ 181
C. Quantitative Analysis .................................. 187
D. References ............................................. 190
E. Sampler Development Committee ....................... 192
The Council of Chief State School Officers

The Council of Chief State School Officers (CCSSO) is a nationwide, nonprofit organization composed of the public officials who head departments of elementary and secondary education in the states, the District of Columbia, the Department of Defense Education Activity, and five extra-state jurisdictions. The Council of Chief State School Officers seeks its members' consensus on major educational issues and expresses their views to civic and professional organizations, federal agencies, Congress, and the public. Through its structure of standing and special committees, the Council responds to a broad range of concerns about education and provides leadership on major education issues.

Because the Council represents the chief education administrators, it has access to the educational and governmental establishment in each state and to the national influence that accompanies this unique position. The Council of Chief State School Officers forms coalitions with many other education organizations and is able to provide leadership for a variety of policy concerns that affect elementary and secondary education. Thus, CCSSO members are able to act cooperatively on matters vital to the education of America's young people.

Council of Chief State School Officers
Peter Mc Walters (Rhode Island), President (2000-2001)
Suellen K. Reed (Indiana), President-Elect
Ted Stilwill (Iowa), Vice President
Gordon M. Ambach, Executive Director
Council of Chief State School Officers
One Massachusetts Avenue, N.W., Suite 700
Washington, DC 20001-1431
202-408-5505
Fax: 202-408-8072
www.ccsso.org
ISBN: #1884037-70-4

The Council's Assessment Center

The Council’s State Education Assessment Center was established to provide an information base on education in the United States, especially from a state perspective. The Center works to improve the breadth, quality, and comparability of data education, including state-by-state achievement data; descriptive data; indicators of quality in areas such as math and science; and performance assessment of students, teachers, and education leaders.

In collaboration with state education agencies, the federal government, and national and international organizations, the Center contributes to a set of useful and valid measures of education geared, when appropriate, to education standards. The Center also supports efforts by states to use standards to improve instruction through collaborative activities of states and others within the education field. Finally, the Center sponsors the Annual Large-Scale Assessment Conference, a meeting devoted to a wide range of topics regarding assessment of students enrolled in public elementary and secondary schools.

The Council’s Resource Center

The Council’s Resource Center on Educational Equity was established by chief state school officers to provide services designed to ensure equitable, high-quality, and developmentally appropriate education for all students, especially minorities, females, students with disabilities, limited English proficient students, and low-income students. The Resource Center conducts research and policy formulation, develops reports and other materials, operates grant and other action programs, provides capacity-building technical assistance to state education agencies, holds working conferences, and monitors federal and state civil rights and education programs that focus on disadvantaged students.
The State Collaborative on Assessment and Student Standards (SCASS) was created in October 1991 to encourage and assist states in working collaboratively on assessment design and development in a variety of subject areas. The State Education Assessment Center of the Council of Chief State School Officers is the organizer, facilitator, and administrator of the projects. A total of 46 states and 2 extra-state jurisdictions participated in nine projects during this project year from July 1999 to June 2000.

States gain most from the primary products of the SCASS projects. All products are determined and designed by the participating states based on their particular needs. Some SCASS projects pool their resources to conduct cutting-edge research (such as the Comprehensive Assessment Systems (CAS), LEP, Surveys of Enacted Curriculum (SEC) and Assessing Special Education SCASS (ASES) while others commission papers or write reports to define, clarify, or interpret assessment-related issues. Some projects develop guides for helping educators understand and use assessments or build training programs for creating and using portfolios. Three projects are currently creating CD-ROMs to package large quantities of assessment items and materials in order to provide them to state personnel, teachers, and other educators in a user-friendly way.

Because these projects are collaborations among collections of states, costs for developing assessment items and implementation materials are shared over the number of states involved (such as the science, health, and social studies projects). In comparison, a single state that may want to contract with a major test publisher to develop a secure test form at a single grade level may require a contract of more than one million dollars. Other benefits of SCASS projects are professional development opportunities for participating teachers from member states and the opportunity to collaborate with the U.S. Department of Education on several assessment-related tasks of interest to the Department.

The SCASS Assessing Limited English Proficient Students Consortium

The Assessing Limited English Proficient (LEP) Students Consortium develops procedures and materials to assist states towards more appropriate assessment of English Language Learners (ELL) students, including research on effective programs for ELL students, language proficiency measures, and other materials related to measuring academic achievement. The project is jointly administered by staff from the Council's Resource Center on Educational Equity and the State Education Assessment Center.

The members, participants, and consultants of the SCASS LEP group are engaged in the development of various products designed to support standard-based assessment for limited English proficient students. Several products have been completed to date which include: A Guide to Scoring LEP Student Responses to Open-Ended Mathematics Items (1998), and A Guide to Scoring LEP Student Responses to Open-Ended Science Items (1999). These guides have been arranged to be adapted in response to various large scale assessment needs. They contain linguistic training guidelines, examples of students' work, a discussion of issues related to the accurate development of assessments appropriately geared toward English language learners, and a glossary of terms. These publications are tools for training scorers of LEP students' responses to open-ended, mathematics and science items. Both can be used as training tools to increase the accuracy with which scorers evaluate work completed by ELL students.
A Conceptual Framework for the Valid and Comparable Measurement of All Students is a paper that presents a conceptual framework and research plan on the topic of test standardization. This document is the outcome of a series of meetings convened during the winter of 1998 with technical experts, educational researchers, student advocates, and curriculum experts on this topic. The purpose of the meetings was to start a discussion regarding changing the theoretical framework used to construct assessments. Participants examined the discrepancies between a "one-size fit-all" approach and how children actually learn, process information, reason and respond in terms of the definition of standardization. The recommendations generated at this meeting and captured in the report will serve as the basis for further research.

In addition to the LEP Student Performance Sampler, which is the contents of this report, the LEP SCASS project staff and consultants have developed a guide entitled: Ensuring Accuracy in Testing for English Language Learners: A Practical Guide for Assessment (2000), which examines the complex issues of testing students with limited proficiency in English and provides help in developing and administering tests for these students. The guide also makes specific recommendations for test producers and administrators.

Introduction: The Challenge of Testing English Language Learners

The adoption of a standards-based educational reform at the national and state levels has resulted in the inclusion of limited English proficient students in large-scale assessment programs at the national, state and district levels. A major challenge in the assessment of these students derives from the fact that academic skills cannot be separated from language skills. Therefore, assessing the content knowledge of LEP students in English becomes an assessment of their English language skills. All types of instruments created with the intent to measure academic achievement, from multiple choice tests to performance assessments, depend on the use of language and are based on assumptions about the students' capabilities to understand written and spoken language and to communicate their ideas in writing (García & Pearson, 1994). This confounding of language comprehension with content knowledge raises questions about the construct validity of the assessments (Durán, 1989).

Prior to the onset of standards-based reform, the solution to this problem was to exclude the students from large-scale assessment for a specified period of time until the students were viewed as sufficiently equipped with the English language skills necessary to participate in the mainstream assessment. As a consequence, little was known about LEP students' knowledge of content during the period of exclusion. Moreover, districts or schools were not held accountable for the education of these students. Standards-based reform has reversed these policies and concomitant beliefs about the exclusion of LEP students. In the past five years, we have seen a shift in state policies and

---

1 The exception was for those students who were receiving content instruction in their native languages.
practices as they relate to the education of LEP students. For example, in a survey conducted by CCSSO in 1995, 15 states included LEP students in their statewide assessment program. In 1999, the number of states that included LEP students in statewide assessments had risen to 48. Policymakers and educators recognize that society cannot afford to neglect an increasing percentage of LEP students in public schools.

State and district officials are working to develop assessment systems for obtaining valid measures of LEP students' knowledge of content. Current research and development efforts will add to the body of knowledge about LEP student assessment and yield appropriate assessment instruments for this population of students. Approaches intended to increase access of LEP students to large-scale tests include translating of tests in English (or portions of them), the use of hands-on tests, and allowing accommodations in standard mainstream tests. Of the approaches listed above, test accommodations seem to be the most prevalent strategy at the state level. The most recent Council report on state assessment practices shows that in 1997-98, 38 states allowed accommodations on presentation format (an increase from 33 states since 1996-97); 37 states allowed setting accommodations (an increase from 31 states); 39 states allowed timing/scheduling accommodations (an increase from 28 states); and response format accommodations increased from 10 to 16 states. However, many questions about the effectiveness of the accommodations to produce valid information on the academic achievement of LEP students remain unanswered.

Accommodations include: changing the presentation of the English language test, its administration, and response options available to students.
The Sampler Project
Part I: The Sampler Project

In an attempt to contribute to the field’s understanding of practices that increase LEP student participation in large scale assessments, the Council of Chief State School Officers has worked on the development of efforts that focus on student’s responses to open-ended assessment items. Important outcomes from those efforts are guides to scoring the responses of LEP students to mathematics and science open-ended items (Kopriva & Saez, 1998; Kopriva & Sexton, 1999). These guides provide raters with basic linguistic information critical to the proper interpretation of English learners’ responses to test items. The discussion is illustrated with student response samples.

We decided to continue these efforts by putting together this sampler of student responses to science and mathematics items. Our goal is to show how student responses differ depending on whether students are tested in English or in their native languages. The sampler intends to promote awareness among teachers and assessment developers that low language proficiency (in English) is a factor that can penalize unfairly the scores of English learners in science and mathematics assessment. Ultimately, the sampler intends to promote awareness that language and culture are complex, interrelated phenomena that influence the way in which students interpret science and mathematics and respond to them (Lee, Fradd, & Sutman, 1995; Solano-Flores & Nelson-Barber, 2000).

This document is comprised of five parts including the work samples and a set of appendices. Part I describes the approach used by the consultants and staff to develop the sampler. This includes the selection of students and items, test administration, scoring of the items, and an explanation of the layout of the sampler. Part II presents a summary discussion of the quantitative and qualitative analysis showing the degree of concurrence in teacher ratings of student responses within the same language and across languages. Part III discusses the implications of the results obtained and offers recommendations. Part IV presents suggested exercises designed to promote the use of the sampler as a professional development tool. Finally, Part V shows the student work samples.

Methods Used to Develop This Sampler

Levels of Analysis

To develop this sampler, we used two levels of analysis, group and individual, respectively supported by a quantitative and a qualitative approach. At the group level, a quantitative approach, allowed us to examine how the scores of students on several items vary depending on the language in which they are tested. It also allowed us to assess whether the lack of consistency between the scores given by independent raters is a factor that may threaten the validity of conclusions based on test scores. At the individual level, a qualitative approach allowed us to examine how proficiency in English, proficiency in the native language, and culture influence the quality of responses from selected students.

Native Languages

Three native languages, Spanish, Chinese, and Haitian-Creole, were selected to be included in this sampler. These languages to some extent reflect the linguistic and cultural diversity of this nation. In addition to being considerably different from each other, they are spoken by very different cultural segments of the population. Also, they are relevant to testing and education for different reasons. For example, Spanish is relevant because it is the native language spoken by the majority of English learners in the United States; it is also the language used in 75 percent of the bilingual programs in this country. By contrast, Haitian-Creole is interesting because only a
small number of linguistic minority students speak it, most of whom are enrolled in Northeastern states and Southern Florida. Consequently, the set of challenges faced by those who teach Haitian-Creole native speakers are different from those who teach Spanish native speakers. Similarly, Chinese students represent a small but growing proportion of the ELL student population. Educators in the community in which these students live can benefit greatly from understanding the nature of these students' responses.

Participants

A total of 124 students participated in this project. Of these students, 39 were Spanish speakers, 62 spoke a Chinese language, and 23 were Haitian-Creole speakers. These students were either first or second generation in the United States, mainly from Mexico and Central America for Spanish, Mainland China and Hong Kong for Chinese, and Haiti for Haitian-Creole.

A total of eight teachers were selected for the project, as instructors of the participating students. The teachers' participation involved translating science and mathematics test items into their respective native languages, administering them to their own students, and scoring and analyzing student responses. Of these eight teachers, three were teaching in bilingual Spanish programs, two were bilingual Chinese teachers, and three taught in bilingual Haitian-Creole classrooms.

These teachers were native speakers of the languages included in the project. In the case of Spanish, one of the three teachers was born in Nicaragua, another in the United States of Mexican-American parents, and the other was born in Mexico and migrated to this country at a young age. Two taught in California, and one in Texas. In the case of Chinese, one teacher was born in Taiwan and the other in Mainland China; both had migrated to this country as adults. One taught in California and the other in Pennsylvania. In the case of Haitian-Creole, the three teachers were born in Haiti, had migrated to the United States as adults, and taught in the state of New York.

Item Selection

We selected four open-ended science items and three open-ended mathematics items from the fourth-grade 1996 set of items released by the National Assessment of Educational Progress [NAEP] (1996a; 1996b) (Table 1). We could use a small number of items because the dependability of total, aggregated scores was not an issue for the purposes of this project. With this small number of items, it is easy for the reader to examine a wide variety of responses. By using this set of NAEP items, which are deemed psychometrically sound, we ensured that item technical properties and the procedures used for item writing and scoring rubric development would not be issues that could threaten the defensibility of the conclusions drawn from this project. This sample was intentional because we wanted to include different areas within science and mathematics.

Item Translation

To address the fact that even expert translation in assessment is extremely sensitive to subtle cultural differences among students from the same linguistic group (Solano-Flores, Trumbull, & Nelson-Barber, 2000), the items were translated by the teachers who taught the students who participated in this project. In addition, to verify the accuracy of the translations, an independent translator subsequently reviewed the translations.

We convened a full-day meeting with the teachers to translate the selected NAEP items. A science assessment specialist with extensive experience in test translation and a skilled specialist in bilingual education conducted this meeting. In the first part of the
We facilitated a group discussion on linguistic issues relevant to testing English learners. In this discussion, the teachers shared with the group their experience teaching and testing LEP students and commented on the challenges of translating those specific NAEP items for their students. Several issues arose. For example, in *Gumball Machine*, students have to figure out how many gumballs of a given color may come from a gumball machine, based on information on the numbers of gumballs for each color that are in it. It turned out that probability is not taught formally in China during the first nine years of instruction; therefore, the Chinese teachers warned us that their students' scores for that item might be lower. Also, one of the Haitian teachers said that this problem might not favor Haitian students who had migrated recently to this country because gumball machines are almost nonexistent in Haiti. These important differences in student experiences and teaching practices emphasize the fact that language and culture cannot be separated.

In the second part of the meeting, the group split into teams to translate the items. Despite the small number of items and the relatively short text used in each, the translation task took several hours because the teachers discussed at length every single detail of the translations. We instructed the teachers that they needed to ensure that the translations were appropriate for their students. Also, the teachers had to reach a consensus on how they would address differences in language use such as regionalisms and frequency of some words. This was particularly the case for the Chinese translation because of subtle but important differences between the Chinese spoken in Hong Kong and the Chinese spoken in Mainland China.

Because the translations were intended to suit the characteristics of their students, we made an effort not to express opinions that might influence the decisions the teachers made during the process of translation. For example, although
the two researchers who facilitated the discussion are native Spanish speakers, they did not get involved in the discussion the Spanish team had about word usage, spelling, or accents.

A few days after the translation session took place, we sent the teachers a "clean" typed version of the translated items for their review. After minor refinements, we produced a test booklet in English for the science items a test booklet in English for the mathematics items, and translations of those booklets in Spanish, Chinese, and Haitian-Creole. These Spanish, Chinese, and Haitian-Creole test booklets were as similar in appearance to the English booklets as possible. Needless to say, this could be done only within certain limits. For example, because of the length of words and their syntactical structure, printed Spanish uses more text than English and the size of the Chinese characters was slightly bigger than the font sizes used in the other language versions. In some test translations, this has the undesired consequence of reducing the space students are given to provide their responses, but this was not the case in our project. Each item was printed on a separate page.

Test Administration

Teachers administered the test to their own students in two testing periods of no longer than 45 minutes each, seven to ten days apart. Since writing in Chinese takes longer than writing in the other languages included in the project, the test administration time was a bit longer for the Chinese students. This way we ensured that writing in Chinese would not interfere with performance. Each student completed four booklets: science in English, science in their native language, mathematics in English, and mathematics in their native language. To control for the effects of learning and practice, each student was assigned randomly to one of the four testing sequences shown in Table 2. Accordingly, students completed booklets on the same subject at least seven days apart, and the two booklets they completed within the same testing period were for different subjects and appeared in different languages.

Table 2. Testing sequences used in the project.

<table>
<thead>
<tr>
<th>Testing Sequence</th>
<th>First Testing Period</th>
<th>Second Testing Period</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>First Booklet</td>
<td>Second Booklet</td>
</tr>
<tr>
<td>1</td>
<td>Mathematics</td>
<td>Science in English</td>
</tr>
<tr>
<td></td>
<td>in native language</td>
<td>native language</td>
</tr>
<tr>
<td>2</td>
<td>Mathematics</td>
<td>Science in English</td>
</tr>
<tr>
<td></td>
<td>in native language</td>
<td>native language</td>
</tr>
<tr>
<td>3</td>
<td>Science in English</td>
<td>Mathematics</td>
</tr>
<tr>
<td></td>
<td>in native language</td>
<td>in English</td>
</tr>
<tr>
<td>4</td>
<td>Science in English</td>
<td>Mathematics</td>
</tr>
<tr>
<td></td>
<td>in native language</td>
<td>in English</td>
</tr>
</tbody>
</table>

Scoring

In the original design of the project we selected three teachers for each of the languages: Spanish, Chinese, and Haitian-Creole. Within each language, every teacher would score the responses of the other two teachers' students, so that every response would be scored by two raters and the teachers would not score the responses of their own students. However, due to the teachers' schedules, this design could not be attained for the Chinese participants. Since only two Chinese teachers were available, we had them score the responses of each other's students as well as their own students' responses. In the case of Haitian-Creole, double-scoring could be completed only for the students of one teacher. As a result, only the responses from students for that teacher were included in the sample.
Also, in the original design we had included two teachers whose experience was limited to teaching mostly Anglo mainstream students, who were native English speakers, and who did not speak any other language. These teachers would double-score the responses to the items in English for portions of the three native language groups. Our intent was to perform a series of quantitative analyses to assess the consistency of their scoring. We suspected that their lack of familiarity with any of the native languages would limit their ability to interpret student responses in English, and that would render lower scoring consistency between raters. Unfortunately, this extra pair of teachers could not come to the scoring session, and we could not obtain this valuable piece of information.

As a result of these changes, we ended up with the scoring design shown in Table 3. Each of the seven NAEP items was scored twice for the English version and twice for any given native language version. Since 124 students participated in the study, the scoring involved 3472 item-scoring events. This scoring was completed during a two-day meeting that we convened with the teachers.

### Table 3. Scoring for any given item for the three native language groups.

<table>
<thead>
<tr>
<th>Spanish Native Speakers</th>
<th>Chinese Native Speakers</th>
<th>Haitian-Creole Native Speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher and Class Size</td>
<td>Teacher and Class Size</td>
<td>Teacher and Class Size</td>
</tr>
<tr>
<td>A (10)</td>
<td>B, C</td>
<td>D (27)</td>
</tr>
<tr>
<td>B (20)</td>
<td>A, C</td>
<td>D, E</td>
</tr>
<tr>
<td>C (9)</td>
<td>A, B</td>
<td>E (35)</td>
</tr>
<tr>
<td>Total N: 39</td>
<td>Total N: 62</td>
<td>Total N: 23</td>
</tr>
</tbody>
</table>

Prior to the training, the teachers were provided with conceptual frameworks and a detailed task analysis of the items, aligned with the National Science Education Standards (National Research Council, 1996), and the standards of the National Council of Teachers of Mathematics (1989). The intent was for them to incorporate in their analysis of student responses a breakdown of the skills and abilities the students could or could not perform. Our purpose was to bring the teachers to the same level of discourse used by NAEP in the content they would score.

Teachers were trained to score the student responses to each item with the scoring rubrics used by NAEP. Based on comments by teachers and to ensure proper interpretation, minor changes in wording and punctuation were made on some scoring rubrics during the training. The sample responses included in the Nation's Report Card The Math and Science Reports (National Assessment of Educational Progress, 1996a, 1996b) were used as anchor responses.

No attempt was made to translate the scoring rubrics into any of the native languages included in this project. The same scoring rubrics in English were used to score both the responses of the students when they were tested in English and when they were tested in their native languages.

A science assessment specialist, a specialist in bilingual education, and a science teacher with extensive experience conducting scoring sessions facilitated the training. For each item, the training involved a group discussion intended to examine the characteristics of the scoring rubrics and the characteristics of the anchor responses. Next, the teachers practiced using the scoring rubrics to score independently randomly selected responses in English from students from the three native language groups. After scoring each response, each teacher shared with the group the score s/he had given and the reasons for the differences were discussed until consensus was reached. The process was repeated until all the teachers assigned the same score to student responses three consecutive times.
Data Collection

In order to have external measures of academic achievement against which we could compare the scores given by our teachers, we attempted to obtain standardized scores for science, mathematics, and reading proficiency. However, we decided not to use this information because it was incomplete.

Nonetheless, we were able to obtain valuable, additional information regarding the students’ abilities and academic achievement. We had the teachers rank-order their students into five levels of proficiency for six areas: proficiency in the native language, literacy in the native language, proficiency in English, literacy in English, and proficiency in science and mathematics. We asked the teachers to provide this information several weeks before the scoring session took place, so that their judgments of their students’ abilities and academic achievement could not bias their scoring decisions. Despite the limitations (e.g., they cannot be used as standardized measures across classrooms), these rank-orderings provide information that the readers of this sampler may find helpful for interpreting the students’ responses.

A third, two full-days meeting was convened with the teachers to analyze the student responses. The goal was to select responses that showed how student responses differ depending on what language is used.

In the first part of this meeting, two specialists in bilingual education gave a presentation on basic concepts in linguistics and discussed with the participants linguistic and cultural issues relevant to learning a second language. This training was intended to help the teachers understand what they should look at when they read the student responses.

In the second part of the meeting, we asked the teachers to work in teams within each native language to compare the responses given by each student on the two language versions and to select those responses that showed important differences across languages. To ensure that the teachers’ evaluation of the students’ responses was not influenced by previous scores, we did not disclose the scores obtained by the students before or during this meeting. Three rounds of selection of responses were completed until the teachers reached a reduced number of responses that best showed performance differences across languages.

In the third part of the meeting, teachers were asked to provide in-depth commentary about the linguistic features reflected in the students’ responses in both the English and native language. They also were asked to interpret any differences or similarities based on their own knowledge of the native languages. In addition, teachers were asked to judge the quality of the students’ responses with respect to understanding the mathematics or science concept or skill assessed by the item. To assist the teachers with the latter task, the scoring rubrics that had been used in the scoring sessions were made available in this meeting. Finally, teachers translated into English each native language samples they had selected.

To ensure conceptual coherence and an adequate level of technical formality, we incorporated the perspective of linguistics into the commentaries written by the teachers. In doing so, we also ensured that the style used in all the commentaries was consistent throughout the entire sampler.
Design of the Sampler

The sampler is organized by content area (science and mathematics). Our intent was to ensure both the flavor and authenticity of the sampler and to provide readers with information they could use to properly interpret the student responses.

Within each content area, student responses are shown by item; within each item, responses are shown by student native language: Spanish, Chinese, and Haitian-Creole. The number of responses included varies depending on the language.

For each student included, the following components are shown (See Figure 1):

1. **Student's name.** A fictitious name used in the sampler to protect the student's identity. At the end of the project, we randomly assigned fictitious first names to the student from a pool of names in the three native languages. Because we are not familiar with Chinese and Haitian-Creole, we may not have always been able to tell girls' names from boys' names. Consequently, in some cases there also may be a mismatch between the fictitious name and the gender implied in the commentaries. Also, we are aware that Hong Kong and Mainland China differ both in writing style and the way in which names are written in Latin characters. Because of our lack of familiarity with Chinese, in some cases there may also be a mismatch between the style in which the fictitious name is written and the student's origin, as reflected by the writing style.

2. **Teacher ratings of the student's language and academic skills relative to their peers.** These teacher ratings are given in a 1 (minimum)-5 (maximum) scale. Ratings on six areas are provided: English fluency, English literacy, fluency in the student's native language, literacy in the student's native language, science proficiency, and mathematics proficiency.

3. **Response in the student's native language.** This is the scanned, reduced image of the item and the student's response in their native language.

4. **Response in English.** This is the scanned, reduced image of the item and the student's response in English.

5. **"Typed" student's response in English.** Although Tadpole and Frog were analyzed as separate items, they are presented together in the sampler.

6. **English language translation of the student's response given in his or her native language.** In analyzing this section, it is important to be aware that as an effect of translation, the quality of the responses of some students in their native languages may look better than the quality of the original response because some mistakes that cannot be preserved in a translation. In order to convey the meaning intended by the student in the native language, translators have to "correct" the language used by the student. For example, a student wrote 'allejas' in the Spanish version of the Tadpole exercise when he actually meant to write 'agallas'. However, the response is translated into English as 'gills', as if the student had written this word correctly in Spanish. The same happens with sentence structure and punctuation. Especially when the reader is not familiar with the student's native language of the response analyzed, it is important to bear in mind that this "inflated" quality of the translated response may in some cases produce the erroneous impression that the commentary or the scores are a bit unfair or not well-focused.
Figure 1. Student response sample layout.

Content Area: Science
Languages: English, Spanish

Teacher Ratings:

<table>
<thead>
<tr>
<th>Item: Tadpole/Frog</th>
<th>Fluency in 1st Language</th>
<th>Literacy in 1st Language</th>
<th>Fluency in 2nd Language</th>
<th>Literacy in 2nd Language</th>
<th>Science Proficiency</th>
<th>Math Proficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Karina</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
</tbody>
</table>

1. Todos los animales necesitan oxígeno para vivir. Explica cómo un renacuajo obtiene oxígeno en su cuerpo. * All animals need oxygen to live. Describe how a tadpole gets oxygen into its body.

El renacuajo respira por aletas. * The tadpole breathes through gills.

2. Explica cómo una rana obtiene oxígeno en su cuerpo. * Describe how a frog gets oxygen into its body.

La rana respira por los pulmones. * The frog breathes through the lungs.

Rating Scale
1. Minimum
2. Maximum

English Response:
I think that the tadpole gets oxygen by its gills.

Translation from Native Language:
I think that the frog gets oxygen by its gills.

CCSSO LEP Work Sampler Booklet 17
**Item Scores**: 0 Minimum to 3 Maximum

<table>
<thead>
<tr>
<th>English Response (Tadpole, Frog)</th>
<th>Native Language Response (Tadpole, Frog)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scorer 1: 1, 2</td>
<td>Scorer 1: 1, 1</td>
</tr>
<tr>
<td>Scorer 2: 1, 2</td>
<td>Scorer 2: 2, 1</td>
</tr>
<tr>
<td>Mean Score: 1.2</td>
<td>Mean Score: 1.5, 1</td>
</tr>
</tbody>
</table>

**Notes**:

7. **Item Scores.** Scores given by each rater in the English version and scores given by each rater in the student's native language. Mean scores obtained by averaging across raters within each language are also provided. Separate scores for Tadpole and Frog are provided on the same page.

8. **Commentary.** Brief analysis of the student's responses. This commentary focuses on the differences in language use across languages and its relation to the quality [understanding of the concepts/skills] of the student's performance.

9. **Notes.** Space for readers to take notes.

---

In the English version, the student identifies lungs as the breathing structure of a frog. Because of its sound in Haitian-Creole, "e" is used in place of "y" in oxygen.

In the Haitian-Creole version, there is a slight indication that the student identifies the source of oxygen.

There is no strong evidence in either language that the student understands the concepts, or that she knows the amphibian's respiratory structures and functions. The major mean score differences across languages is for Frog (2 and 1, respectively in English and Haitian-Creole) due to the fact that knowledge of the frog's breathing organs is not demonstrated consistently across languages.
We performed two kinds of analyses, quantitative and qualitative. Both levels of analysis are important. The former provides a broad picture of testing English language learners as a group. It can help policymakers and other decision makers to design more effective testing strategies for linguistic minorities. The quantitative analyses provide concrete examples of how the characteristics of the responses of specific students vary across languages. It can help teachers and practitioners to reflect about the complex relationship between language and testing.

**Quantitative Analysis**

The quantitative analysis focused on score differences across languages. This analysis was performed separately for each native language group and included all students who participated in the study. We examined whether testing English language learners in their native language works always to their benefit. If this is entirely true, then we should expect that all of these students consistently perform better on all items when they are tested in their native language than when they are tested in English.

As a part of the quantitative analysis, we also examined whether judgement inconsistencies across raters and other sources of measurement error could affect the dependability of the scores.

A technical discussion of the quantitative analyses and results is provided in Appendix C. The results can be summarized as follows. First, the scoring of independent raters was reasonably consistent; it did not have an adverse impact on the reliability of the scores.

Second, students did not perform consistently better when they were tested in English or in their native languages. These results are shown in Figure 2. The bar graph shows the difference between mean scores on each item when students were tested in their native languages and when they were tested in English. For each group, a positive (upward) direction of the bar indicates a difference in favor of the native language and a negative (downward) direction of the bar indicates a difference in favor of English. The graph shows that the students performed better in their native languages for some items and in English for other items.

A more sophisticated analysis based on generalizability theory (see Appendix C), rendered consistent results. The main source of score variability was due to the interaction of student, language, and item.

**Figure 2. Difference between mean scores obtained by the students on each item when they were tested in English and when they were tested in their native languages.**
Qualitative Analysis

The qualitative analysis focused on differences and similarities in the student responses across languages. This analysis was performed for each of the student responses included in this sampler.

We examined the strengths and weaknesses of the student responses in the English and native language versions of the test. The commentary provided for each sample discusses the student responses based on accuracy (i.e., correctness), technical sophistication (e.g., proper use of scientific or mathematical terms), and writing proficiency (e.g., sentence structure and clarity).

The student response samples show that English language learners do not necessarily perform better when they are tested in their native languages, which is consistent with the results from the quantitative analysis. The scores are determined by the complex interaction of the characteristics of the students, the characteristics and content of each item, and the language of the test.

Table 4 illustrates this interaction among student, item, and test language. It shows the responses given and scores obtained by three specific students on the science items in both English and their native languages. This table shows the pattern most typically observed. First, the quality of the responses varies with item. Students perform well on some items and poorly on others. Second, the quality of the responses varies also with language. For some items, students perform better in their native languages than in English. For other items, their performance is better in English than in their native languages. (Of course, there are some students who consistently perform better in English or their native languages, or students whose performance is consistently high or consistently low in both languages. But these cases do not constitute the majority.)

<table>
<thead>
<tr>
<th>Item</th>
<th>Response in English</th>
<th>Score</th>
<th>Response in native language (translated)</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tadpole I thin [...]</td>
<td>.5</td>
<td>A tadpole gets oxygen in its gills.</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Frog A frog gets oxygen into his body in his mouth too.</td>
<td>1.5</td>
<td>A frog gets oxygen in its lungs.</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Erosion [Picture B]. Because their more mountains and it makes more water.</td>
<td>1</td>
<td>[Picture B]. Because the river of today has more mountains and it makes more water.</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Heat The nails become warm because if you hit it into a piece of wood when it goes all the way down its going to become warm.</td>
<td>2</td>
<td>The nail gets warmer when somebody hammers it on a piece of wood all the way in.</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Tadpole Maybe it Breathe but you can't see that it breathe and maybe it breathe in it's body.</td>
<td>1</td>
<td>It takes oxygen in its mouth because it lives in water.</td>
<td>1.5</td>
<td></td>
</tr>
<tr>
<td>Frog Maybe it's lungs grab oxegen But not through it's nose. But from another hole.</td>
<td>1</td>
<td>It takes oxygen in its nose because it lives on earth.</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Erosion [Picture A]. I pick A because mountains are not cacky and how mountains are open to let the River go through.</td>
<td>2</td>
<td>[Picture B]. I choose B because the mountains do not look smooth.</td>
<td>1.5</td>
<td></td>
</tr>
<tr>
<td>Heat Because the wood is warm and it is a good insolator.</td>
<td>1.5</td>
<td>The nail becomes hot because the wood is hot and you are hitting the nail.</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>
Part III: Implications and Recommendations

The results show that reasonable interrater reliabilities can be obtained both when students are tested in English and when they are tested in their native languages. However, this conclusion currently applies only if the raters are native speakers of those native languages. Further research needs to be done to examine whether comparable results can be obtained when the raters are monolingual, native English speakers.

The results also indicate that student performance varies considerably with the language used to test them. Testing English language learners in their native language may or may not work in their favor depending on a number of factors. Students may perform better in their native language than in English for some items but worse in their native language than in English for other items. As many of the responses included in this sampler illustrate, this is especially the case for students who are not very proficient in their own native languages.

Whereas we agree with the notion that assessing English learners in their native languages can serve as a way to measure the continued development of their content knowledge—especially when they receive instruction in their native languages (Stansfield, 1996), we think that assuming that these students should be tested only in their native languages would be too simplistic.

As with many issues in bilingual education, no simple rules can be formulated to ensure more dependable scores and more equitable testing for English learners, especially those who are not proficient in their native languages. However, based on both the results reported above and the experience gained from developing this sampler, we can make three recommendations for assessment developers, teachers, and decision makers.
Recommendation one:
- **Include bilingual education and linguistics specialists as members of the assessment development teams.** These specialists should be involved at all times in both the development of the assessment prompts and the development of the scoring rubrics. Their expertise should help the teams to identify and to address properly cultural factors that are critical to assessment across languages. There is evidence that even when assessment developers are bilingual, they cannot take into consideration subtle variations in language use and certain cultural influences that are critical to properly interpreting assessment prompts. As a consequence, the assessment given to LEP students in their native language may reflect the assessment developers’ thinking rather than the students’ thinking (Solano-Flores, Trumbull, & Nelson-Barber, 2000).

Recommendation two:
- **Ensure that scoring training sessions incorporate training in linguistic issues relevant to assessment.** Being a native speaker of a language that is not English or belonging to a cultural minority does not make anyone more or less knowledgeable of English language development. Therefore, ensuring the participation of bilingual raters in scoring sessions may not suffice to eliminate cultural or linguistic bias in large-scale testing.

Based on the observed impact that this project had on the participating teachers, we know that a discussion of issues related to assessment across languages and cultures and assessment of LEP students can sensitize teachers to the complex relation of language and culture as well as the influence of this interaction on assessment. If this applies to the participating teachers in this project, who were bilingual, native speakers of languages other than English, and taught in bilingual programs, it should also apply to teachers who are monolingual, native English speakers, and do not teach in bilingual programs.

Recommendation three:
- **Test or pilot-test LEP students in both English and their first languages.** It is well known that, to obtain accurate, reliable and valid scores, students must be tested with a wide variety of exercises a reasonable number of times (Othman, 1995; Ruiz-Primo, Baxter, & Shavelson, 1993). This principle should be extended to the language that is used in testing. To ensure equitable testing and more accurate measures of their achievement, LEP learners should be tested with a wide variety of exercises in both their first and second languages. This does not mean necessarily that these students should always be given bilingual, side-by-side tests. Nor does it mean that students always should be tested twice. In the context of classroom testing, it means that some items should be administered in English and others in the students’ native languages. In addition to promoting English acquisition, this strategy reinforces the use of the students’ native languages. It also promotes the use of scientific and mathematical thinking in both languages. Moreover, by using this strategy, teachers can obtain more accurate information on both the students’ performance and their command of English and their native languages.

In the context of large-scale testing, the recommendation is based on the fact that the development of an assessment must be cyclical. In this cyclical process, the task, the scoring system, and the response format (especially its wording) must be continuously refined based on the responses given by students (Solano-Flores & Shavelson, 1997). During this process, when draft versions of an assessment are tried out with pilot students, LEP students should be pilot tested in both English and their native languages. By doing so, assessment developers can gain
more knowledge on how students interpret items and what kind of thinking the items elicit in each language.

This strategy can render information that is critical to item wording. For example, based on the students' responses in both languages, assessment developers can become aware of words in English which students use because of the existence of cognate words in their native language (words that are similar in both languages because they have the same root). In some cases, using such cognates can be to the advantage of English learners, as when they use *respiration* in English because it is similar to *respiración* in Spanish. In some other cases, it can be to their disadvantage, as when they use *lemon* in English because it is similar to *limón* (lime) in Spanish.

The development of the scoring rubrics can also benefit from this strategy. For example, from examining the students' responses in both English and Haitian-Creole, the Haitian teachers who participated in this project realized that *bouth* results from mixing *mouth* in English and *bouch* in Haitian-Creole, as some responses included in the sampler show. Scoring rubrics could be enriched by including information intended to alert raters about words like *bouth* which at first glance may not make any sense to them and may mislead their scoring decisions. Whereas it cannot be expected that raters should be provided with long lists of words that look senseless along with their intended meanings, some use of this practice can contribute to help assessment developers and raters incorporate a linguistic perspective in their analyses of student responses.

In sum, both classroom and large-scale testing practices can be improved to address the fact that, in addition to prior content knowledge and opportunities to learn, the language and type of item influence how students perform at different levels of achievement. Both the quantitative data and the responses included in this sampler provide evidence that LEP students show different sets of strengths and weaknesses in science and mathematics depending on the language in which they are tested and knowledge of content. Incorporating expertise in linguistics in the process of assessment development, enriching scoring training with information on second language development critical to assessment, and pilot-testing and testing students in both English and their native language seem to be the best approach to address this performance instability.
Part IV: Suggestions for Using Work Samples

The worksamples included in this sampler illustrate how the performance of English language learners on open-ended mathematics and science assessments vary depending on language used to test them. The exercises included in this section intend to help facilitate the planning and implementation of professional development for teachers and other educators.

The exercises are presented sequentially from activities that build awareness of the need for assessments that are sensitive to the developing communication skills of LEP students to scoring the performances of students from three different native language groups. We encourage planners of professional development to promote more in-depth study into the testing of LEP students and to use the student samplers as part of their training programs.

1. **Discuss the influence of English proficiency on testing results.**

   As a warm-up activity, ask participants to work in teams of four and discuss the implications of the following quote.
   
   "Any test is to some degree a test of English language proficiency and may be invalid an unreliable measure of English language learners' academic proficiencies". (American Educational Research Association, 1985).

   After their discussion, ask participants to share the key points and record them on chart paper or a transparency. The information then can be used to remind the participants about the importance of creating valid and reliable assessments for LEP students. As an extension, ask participants to share a successful classroom assessment story they or someone else used that took into consideration the language proficiency of their students.

List the attributes that made the assessment successful to help participants see the variety of ways in which assessments can be designed to be sensitive to LEP students.

2. **Build awareness of the influence of language and content knowledge on students' ability to respond to open-ended assessment items.**

   In the students' work samples the teacher rating is located in the top left-hand corner of the first page of each student sample. The range of the scales is from 1 (minimum) to 5 (maximum). The teachers in the study used their knowledge of the students' language proficiency and mastery of content to rate each student.

   Begin by reviewing the language fluency and literacy scales shown in Appendix A. Hand out one student sample in either mathematics or science to each participant (first page only). Assign participants to work in teams of three or four and direct them to their student samplers based on the following questions:
   
   - How do the ratings help to understand the student's response in English and in the native language?
   - What insights did you get about the student's responses using the rating scales?
   - What other information might you use to help understand the students' response to the assessment item? Why?

   Ask participants to record their responses on a chart paper and display their finished product for the whole group to review. After the teams finish their discussions, ask each group to report their responses to the three questions.

3. **Identify culture and language influences in students' written performance.**

   Discuss with participants the implications of language and culture in assessment. Start by asking participants to recall a time when they or someone they know may have encountered
a situation in which language and/or culture came into play during the assessment of a student. Have participants share their stories and describe how they handled the situation. Next, review the types of culture and language patterns found in student performances in A Guide to Scoring LEP Student Responses to Open-Ended Mathematics Items (1998) and A Guide to Scoring LEP Student Responses to Open-Ended Science Items (1999). Ask participants to work in small groups and identify the culture and language influences in preselected student samplers. To assist participants in examining how different culture and language patterns may influence student performance on different assessment items, use the following chart to select the samples from the three groups used in the study.

<table>
<thead>
<tr>
<th>Open-Ended Assessment Item</th>
<th>English/Spanish</th>
<th>English/Chinese</th>
<th>English/Haitian Creole</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tadpole/Frog</td>
<td>pp. 30 - 30</td>
<td>pp. 42 - 51</td>
<td>pp. 54 - 63</td>
</tr>
<tr>
<td>Erosion</td>
<td>pp. 66 - 73</td>
<td>pp. 76 - 81</td>
<td>pp. 84 - 85</td>
</tr>
<tr>
<td>Heat</td>
<td>pp. 88 - 95</td>
<td>pp. 98 - 103</td>
<td>pp. 106 - 109</td>
</tr>
<tr>
<td>Figures</td>
<td>pp. 112 - 119</td>
<td>pp. 122 - 125</td>
<td>pp. 128 - 135</td>
</tr>
<tr>
<td>Gumball Machine</td>
<td>pp. 138 - 141</td>
<td>pp. 144 - 151</td>
<td>pp. 154 - 155</td>
</tr>
<tr>
<td>Caterpillar</td>
<td>pp. 158 - 159</td>
<td>pp. 162 - 171</td>
<td>pp. 174 - 175</td>
</tr>
</tbody>
</table>

Ask the following questions to help participants examine the student samples.

- What culture and language patterns did you find that were unique in the English and Spanish student samplers? English and Chinese student samplers? English and Haitian Creole student samplers? Did anyone find influences that were common across all three groups?
- How can information on culture and language influences be used to make decisions about student assessment?

- How might the assessment items be improved or revised to help gain a clearer picture of student ability regardless of their culture and language? As time permits, introduce the use of performance tasks and other tasks that do not rely solely on a student's written performance.

4. Score student responses with NAEP scoring rubrics.
For this exercise, make copies of ten student samplers from one test item (first page only) in mathematics or science and the NAEP scoring rubrics in Appendix (B). The following chart shows the number of student samples available by item. Consider making duplicate copies if needed for large groups.

<table>
<thead>
<tr>
<th>Science Samplers</th>
<th>Mathematics Samplers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tadpole/Frog: 15</td>
<td>Figures: 10</td>
</tr>
<tr>
<td>Erosion: 8</td>
<td>Gumball Machine: 7</td>
</tr>
<tr>
<td>Heat: 9</td>
<td>Caterpillar: 7</td>
</tr>
</tbody>
</table>

Once the materials are duplicated, assign participants to small groups and ask each group to read the item and the student responses in English and the native language. If participants are not literate in the native language, have them use the translated version provided in the column on the right hand side of the sampler. Using the NAEP scoring rubric, ask participants to sort their sets of student samples into what they consider to be an appropriate level of proficiency for each language. As participants sort the samples, ask them to be prepared to respond to the following questions.

- Which indicators in the student work did you use for the majority of your work to help sort the student samples?
- Did anyone have problems placing a student sample into one stack over another?
What other criteria might be added to the rubric to consider properly the developing communications skills of LEP students.

- Overall, what did you learn from this sorting exercise?
  Provide time after the sorting exercise for participants to discuss their responses to the questions.

5. Design analytic scoring rubrics to assess student work samples.

As an extension to Activity #4, introduce participants to analytic trait scoring. Judith Arter and Robert Blum (1996) define analytic trait scoring as a scoring procedure in which performances are evaluated for selected dimensions or traits with each trait receiving a separate score. For example, a piece of writing may be evaluated according to organization, use of details, attention to audience, and language usage/mechanics. Working with the entire group, ask participants to brainstorm and identify key indicators or traits from which to score the student samples. Examples of traits the group could identify include:
- Knowledge of basic scientific concepts;
- Accuracy in computing;
- Skills on inferring relationships and drawing conclusions; and
- Ability to communicate clearly and articulately.

After brainstorming, return participants to small groups and have them select three or more traits from which they will develop an analytic scoring rubric. Let the groups know that it is helpful first to determine the highest level of acceptable performances and the lowest level of unacceptable performances for each trait.

The second step is to develop the degrees of development (or adequacy and inadequacy) between the highest and lowest performance points. According to Grant Wiggins (1996) the best descriptors to use are qualitative words or phrases rather than comparative terminology (e.g., “average,” “generally,” or “less accurate”).

As a final step, have the participants test their rubric by scoring student samples to represent each point on their scales. Allow time for participants to modify their rubrics if needed.

6. Revise NAEP rubrics.

As an alternative, have participants revise the NAEP scoring rubrics, based on what they have learned about how language proficiency can affect the scorers' perception of the students' content knowledge. Then, ask participants questions such as:
- What do the scoring rubrics assume in terms of language skills?
- How might those scoring rubrics penalize LEP students?
- How can they be improved?

7. Conduct a qualitative analysis of the student samplers.

Read aloud the commentary of a mathematics exercise from one of the student samplers. The commentaries are located on the second page of the student samples. Ask participants to review the piece more carefully in small groups and identify the strengths and key qualities of the comments that help them to understand the student's performance in English and native language.

After the review, ask participants to share their findings and record their responses on a blank transparency or chart paper. Characteristics the group may identify include:
- Literacy skills in English and native language;
- Knowledge of the content area;
- Second language acquisition; and
- Ability to analyze and organize thoughts.
Next, ask participants to continue working in small groups and review a commentary from a student sample in science. Have the groups again identify the key characteristics used to describe the student’s performance. When the group has completed its task, ask each group to compare the characteristics found in mathematics with those found in science. To prompt the discussions, ask the following questions:

- What characteristics did you find that are similar between the mathematics and science comments?
- What insights did you gain about the student’s performance from the second review of the comments?
- What questions still linger about assessing the performance of English language learners after this exercise?

As an extension, ask participants to take one of the student samples (first page only) and write their own comments about the students’ performance. Encourage participants to use the information they acquired from the earlier exercises to describe the performance of students in English and the primary language. After the exercise, convene the large group and ask participants to share their comments.
Science

Spanish

Item: Tadpole/Frog
### Elena

**Item:** Tadpole/Frog

**Teacher Ratings:***

<table>
<thead>
<tr>
<th>Fluency in 1st Language</th>
<th>Literacy in 1st Language</th>
<th>Fluency in 2nd Language</th>
<th>Literacy in 2nd Language</th>
<th>Science Proficiency</th>
<th>Math Proficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

1. Todos los animales necesitan oxígeno para vivir.

Explica cómo un renacuajo obtiene oxígeno en su cuerpo.

**Translation from English Response:**

I think that the tadpole gets oxygen by his gills.

**Native Language:**

El renacuajo respira por aletas.

2. Explica cómo una rana obtiene oxígeno en su cuerpo.

**Translation from English Response:**

I think that the frog gets oxygen by gills.

**Native Language:**

La rana respira por los pulmones.

---

* Rating Scale
1. Minimum
2. Minimum
to
3. Maximum

---

All animals need oxygen to live. Describe how a tadpole gets oxygen into its body.

**English Response:**

I think that the tadpole gets oxygen by his gills.

**Translation from Native Language:**

The tadpole breathes through gills.

Describe how a frog gets oxygen into its body.

**English Response:**

I think that the frog gets oxygen by gills.

**Translation from Native Language:**

The frog breathes through the lungs.
### Item Scores*:

<table>
<thead>
<tr>
<th>English Response (Tadpole, Frog)</th>
<th>Native Language Response (Tadpole, Frog)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scorer 1: 3, 2</td>
<td>Scorer 1: 2, 2</td>
</tr>
<tr>
<td>Scorer 2: 2, 2</td>
<td>Scorer 2: 2, 2</td>
</tr>
<tr>
<td>Mean Score: 2.5, 2</td>
<td>Mean Score: 2, 2</td>
</tr>
</tbody>
</table>

* Item Scores: 0 Minimum to 3 Maximum

### Commentary:

In the English response, the student misuses 'gils' instead of 'lungs'. In this, her second language, her vocabulary is limited and fails to identify the proper breathing structure.

Although the responses in both languages fail to identify the source of oxygen, the Spanish response is more precise and reflects a better understanding of the breathing structures at different amphibian development stages.

However, scorer one scores the English response for Tadpole a bit too high.
All animals need oxygen to live.
Describe how a tadpole gets oxygen into its body.

**English Response:**

I thin...

**Translation from Native Language:**

A tadpole gets oxygen in its gills.

Describe how a frog gets oxygen into its body.

**English Response:**

A frog gets oxygen into his body in his mouth too.

**Translation from Native Language:**

A frog gets oxygen in its lungs.
**Item Scores**:  

<table>
<thead>
<tr>
<th>English Response (Tadpole, Frog)</th>
<th>Native Language Response (Tadpole, Frog)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scorer 1: 1, 2</td>
<td>Scorer 1: 3, 3</td>
</tr>
<tr>
<td>Scorer 2: 0, 1</td>
<td>Scorer 2: 3, 3</td>
</tr>
<tr>
<td>Mean Score: .5, 1.5</td>
<td>Mean Score: 3, 3</td>
</tr>
</tbody>
</table>

* Item Scores: 0 Minimum to 3 Maximum

**Commentary:**

The student's response in English is incomplete. By contrast, in Spanish the student correctly names the breathing structures of a tadpole and a frog. Yet he still does not identify the source of oxygen.

Not only is the response in Spanish scored higher than the response in English, the scoring of the response in Spanish is more consistent than the response in English. The score '1' given by scorer 2 to the response to "Tadpole in English" ('I thin') may be due to the fact that the scorer was able to read some partial writing from the erased response.
All animals need oxygen to live. Describe how a tadpole gets oxygen into its body. A tadpole breathes with its gills.

It gets oxygen from his gills and lungs. A frog breathes with its lungs.

It gets oxygen from his gills, lungs and he could walk in land and breathe.

1. All animals need oxygen to live.

A tadpole breathes with its gills.

A frog breathes with its lungs.

Translation from English Response:

It gets oxygen from his gills and lungs.

Translation from Native Language:

Translation from English Response:

It gets oxygen from his gills and lungs.

Translation from Native Language:

It gets oxygen from his gills, lungs and he could walk in land and breathe.
**Item Scores**: 0 Minimum to 3 Maximum

<table>
<thead>
<tr>
<th>English Response (Tadpole, Frog)</th>
<th>Native Language Response (Tadpole, Frog)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scorer 1: 3, 3</td>
<td>Scorer 1: 2, 2</td>
</tr>
<tr>
<td>Scorer 2: 2, 2</td>
<td>Scorer 2: 2, 2</td>
</tr>
<tr>
<td>Mean Score: 2.5, 2.5</td>
<td>Mean Score: 2, 2</td>
</tr>
</tbody>
</table>

**Commentary:**

Unlike the English response, the student’s response in Spanish is succinct and accurate, identifying the proper structures for each stage of development of tadpoles and frogs. The student demonstrates a greater command of the Spanish language in response to a specific question and stays on task. This notwithstanding, the raters gave consistently higher scores to the responses in English.
### Teacher Ratings:

<table>
<thead>
<tr>
<th>Fluency in 1st Language</th>
<th>Literacy in 1st Language</th>
<th>Fluency in 2nd Language</th>
<th>Literacy in 2nd Language</th>
<th>Science Proficiency</th>
<th>Math Proficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

### English Response:

All animals need oxygen to live. Describe how a tadpole gets oxygen into its body.

**English Response:**

He goes on top of the water gets a lot of air then he goes back down into the water and that is how he get oxygen and I think from the ears.

### Translation from Native Language:

The tadpole breathes through his gills.

### English Response:

Describe how a frog gets oxygen into its body.

**English Response:**

The frog gets oxygen because of the air and of his nose.

### Translation from Native Language:

The frog breathes through his lungs.
Item Scores:

<table>
<thead>
<tr>
<th>English Response</th>
<th>Native Language Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Tadpole, Frog)</td>
<td>(Tadpole, Frog)</td>
</tr>
<tr>
<td>Scorer 1: 2, 2</td>
<td>Scorer 1: 2, 2</td>
</tr>
<tr>
<td>Scorer 2: 2, 2</td>
<td>Scorer 2: 2, 2</td>
</tr>
<tr>
<td>Mean Score: 2, 2</td>
<td>Mean Score: 2, 2</td>
</tr>
</tbody>
</table>

* Item Scores: 0 Minimum to 3 Maximum

Commentary:

The English response is erroneous, lengthy, and not clearly focused. The Spanish response shows knowledge of the breathing structures of tadpoles and frogs, although it does not make any reference to the source of oxygen. Syntax and spelling in Spanish is also developing. For example, the student writes "ollajas" instead of "agallas."

In spite of the fact that the Spanish responses are more concise and accurate than the English responses, the responses in both languages obtain the same scores, and there is full agreement between scores for both languages.
All animals need oxygen to live. Describe how a tadpole gets oxygen into its body.

**English Response:**

The tadpole gets oxygen in their body because in the water there is oxygen on it and it have gills to breath.

**Translation from Native Language:**

In the water there is oxygen they have gills to breathe. Later the air goes in and with the gills it lets the water out.

Describe how a frog gets oxygen into its body.

**English Response:**

The frogs are alike, like the tadpole in the air there is oxygen they are like us. They have a nose to breath like us. They breath in and out.

**Translation from Native Language:**

They breathe with their nose like us. In the air there is oxygen to breathe. They live in the forest and with that the trees help them.
**Item Scores***:  

<table>
<thead>
<tr>
<th>English Response</th>
<th>Native Language Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Tadpole, Frog)</td>
<td>(Tadpole, Frog)</td>
</tr>
<tr>
<td>Scorer 1: 3, 2</td>
<td>Scorer 1: 3, 3</td>
</tr>
<tr>
<td>Scorer 2: 3, 2</td>
<td>Scorer 2: 3, 2</td>
</tr>
<tr>
<td>Mean Score: 3, 2</td>
<td>Mean Score: 3, 2.5</td>
</tr>
</tbody>
</table>

* **Item Scores**: 0 Minimum to 3 Maximum

**Commentary:**

The student is able to express appropriate responses in both languages, though he fails to recognize the frog’s lungs as its breathing structure. The Spanish response is more elaborate. It states that oxygen in the water is retained while the water is expelled through the gills.

In both languages, the student attempts to elaborate his answers despite vocabulary limitations. (For example, he compares humans’ and frogs’ breathing structures and identifies forests as a habitat for frogs.)

The scores and the agreement between scorers are similar across languages.
Science
Chinese

Item: Tadpole/Frog
Science
Chinese
Item: Tadpole
1. All animals need oxygen to live. Describe how a tadpole gets oxygen into its body.

English Response: The tadpole gets water in its mouth and the tadpole’s gills let the water out and leave the oxygen inside the mouth.

Translation from Native Language: The described gills breathe air then let out the water.

Describe how a frog gets oxygen into its body.

English Response: A frog breathes with their skin, nose or mouth and the oxygen goes in it’s body.

Translation from Native Language: Frogs breathe air which falls into the lungs.
**Item Scores**:  

<table>
<thead>
<tr>
<th>English Response (Tadpole, Frog)</th>
<th>Native Language Response (Tadpole, Frog)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scorer 1: 3, 1</td>
<td>Scorer 1: 2, 2</td>
</tr>
<tr>
<td>Scorer 2: 3, 2</td>
<td>Scorer 2: 2, 3</td>
</tr>
<tr>
<td>Mean Score: 3, 1.5</td>
<td>Mean Score: 2, 2.5</td>
</tr>
</tbody>
</table>

* **Item Scores**: 0 Minimum to 3 Maximum

**Commentary:**

The student's response to the first prompt is correct and is communicated with greater eloquence than in Chinese. There is Cantonese influence in the sentence structure. The native language response shows some misuse of Chinese words. However, the student conveys the same understanding about a tadpole's breathing system in both languages.

The second response to the second prompt is correct in the Chinese language identifying the lungs as the frog's breathing organs. In contrast, the English version lacks specificity and does not identify the lungs properly.

Perhaps due to the influence of the first language, there is more score variability for both the English and the Chinese responses to Frog. Overall, however, the combined score of Tadpole and Frog is the same across languages.
### Zhai Kang

**All animals need oxygen to live.**

Describe how a tadpole gets oxygen into its body.

**English Response:**

Tadpole is the baby frog. It only has a big tail and swim in the pond. First tadpole will take oxygen from air and put oxygen in their mouth. Slowly the oxygen will turn into their body.

**Translation from Native Language:**

When frogs are just born, they are called tadpoles. They have a black, big tail and live in water. They need oxygen and get it into their bodies through their mouths to their intestines.

Describe how a frog gets oxygen into its body.

**English Response:**

When tadpole grows big, it became a frog. A frog has four legs. When oxygen in its mouth, frog mouth will be big and big then oxygen into their body.

**Translation from Native Language:**

The grown-up tadpoles are frogs who have 4 legs. Frogs need oxygen too. When oxygen goes through their mouths, their mouths become bigger and then the oxygen slowly flows into their bodies.

---

**Teacher Ratings**

<table>
<thead>
<tr>
<th>Fluency in 1st Language</th>
<th>Literacy in 1st Language</th>
<th>Fluency in 2nd Language</th>
<th>Literacy in 2nd Language</th>
<th>Science Proficiency</th>
<th>Math Proficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

*Rating Scale*

1. Minimum
2. Maximum
### Item Scores:

<table>
<thead>
<tr>
<th>English Response (Tadpole, Frog)</th>
<th>Native Language Response (Tadpole, Frog)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scorer 1: 1, 1</td>
<td>Scorer 1: 2, 1</td>
</tr>
<tr>
<td>Scorer 2: 1, 1</td>
<td>Scorer 2: 2, 1</td>
</tr>
<tr>
<td>Mean Score: 1, 1</td>
<td>Mean Score: 2, 1</td>
</tr>
</tbody>
</table>

**Item Scores**: 0 Minimum to 3 Maximum

### Commentary:

The student shows limited understanding in both languages, identifying the life cycle of a frog and restating part of the prompt. There are similar explanations in both languages. However, the English version includes the source of oxygen for tadpoles. The Chinese version shows evidence of a larger vocabulary and more mature skill with the language. This is reflected in a higher score for Frogs in Chinese. There is no evidence in either language that the student understood the concept of respiration at the different stages of the frog's life cycle, nor the organs used at each stage of development.
All animals need oxygen to live. Describe how a tadpole gets oxygen into its body.

English Response: A tadpole gets oxygen in the water or in the water into its body. Translation from Native Language: Tadpoles breathe tiny bubbles of oxygen in water or breathe oxygen above the surface of the water.

Describe how a frog gets oxygen into its body.

English Response: A frog use its nose gets oxygen into its body. but most frogs is when the sky dark to get oxygen. Translation from Native Language: Frogs are usually found in fields where there are lots of mosquitos, using their tongues to catch the Mosquitos.
### Item Scores*

<table>
<thead>
<tr>
<th></th>
<th>English Response (Tadpole, Frog)</th>
<th>Native Language Response (Tadpole, Frog)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scorer 1:</td>
<td>2, 1</td>
<td>1, 2</td>
</tr>
<tr>
<td>Scorer 2:</td>
<td>2, 2</td>
<td>2, 2</td>
</tr>
<tr>
<td>Mean Score:</td>
<td>2, 1.5</td>
<td>1.5, 2</td>
</tr>
</tbody>
</table>

* Item Scores: 0 Minimum to 3 Maximum

### Commentary:

The Chinese version shows more articulate explanations, though limited and erroneous in content. The student has a basic idea of the source of oxygen, but does not show understanding of the structure and function of the respiratory system in tadpoles and frogs.

The English version identifies the source of oxygen but has misconceptions, which is evidence of limited understanding of the prompt and the concepts being addressed.

Overall, the scores obtained by the student in Tadpole and Frog balance out across languages.
All animals need oxygen to live. Describe how a tadpole gets oxygen into its body.

English Response:
The tadpole gets oxygen from the water. So then they gets the oxygen to their bodys.

Translation from Native Language:
There is oxygen in water. So tadpoles get oxygen from water. They use their mouths to get oxygen.

Describe how a frog gets oxygen into its body.

English Response:
The frog can gets from land on water. because they are had oxygen. So they gets oxygen into their body.

Translation from Native Language:
Frogs can get oxygen from the shore or on land. They use their belly to get oxygen.
**Item Scores**: 0 Minimum to 3 Maximum

<table>
<thead>
<tr>
<th>English Response (Tadpole, Frog)</th>
<th>Native Language Response (Tadpole, Frog)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scorer 1:</td>
<td>Scorer 1:</td>
</tr>
<tr>
<td>2, 2</td>
<td>2, 2</td>
</tr>
<tr>
<td>Scorer 2:</td>
<td>Scorer 2:</td>
</tr>
<tr>
<td>2, 1</td>
<td>2, 2</td>
</tr>
<tr>
<td>Mean Score: 2, 1.5</td>
<td>Mean Score: 2, 2</td>
</tr>
</tbody>
</table>

**Commentary**:

The Chinese version was longer and more detailed; however, the structure had English language influence and misconceptions.

The student's responses appear to be equally limited in both languages. Although identification of oxygen sources is made, no accurate response is evident on the breathing structure of either organism.
### Teacher Ratings*

<table>
<thead>
<tr>
<th>Fluency in 1st Language</th>
<th>Literacy in 1st Language</th>
<th>Fluency in 2nd Language</th>
<th>Literacy in 2nd Language</th>
<th>Science Proficiency</th>
<th>Math Proficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>

### Jia Xin

All animals need oxygen to live.
Describe how a tadpole gets oxygen into its body.

**English Response:**

A tadpole gets oxygen from its gill and into its body.

**Translation from Native Language:**

Tadpoles use gills to breathe.

Describe how a frog gets oxygen into its body.

**English Response:**

A frog gets oxygen from its lungs like humans. When they open their mouth the oxygen goes in to body.

**Translation from Native Language:**

Frogs use lungs just like people.
### Item Scores*

<table>
<thead>
<tr>
<th>English Response (Tadpole, Frog)</th>
<th>Native Language Response (Tadpole, Frog)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scorer 1: 2, 3</td>
<td>Scorer 1: 2, 2</td>
</tr>
<tr>
<td>Scorer 2: 2, 2</td>
<td>Scorer 2: 2, 2</td>
</tr>
<tr>
<td>Mean Score: 2, 2.5</td>
<td>Mean Score: 2, 2</td>
</tr>
</tbody>
</table>

* Item Scores: 0 Minimum to 3 Maximum

### Commentary:

Although, the English version is more detailed, and reflects both a greater command of the language and use of proper vocabulary, these differences do not affect how the student demonstrates knowledge. Her score on Frog is only slightly higher for English than for Chinese.

The Chinese version conveys the proper content, though it has the influence of English language sentence structure. The student uses Chinese characters to respond, but with inaccurate Chinese syntax and spelling.

The student's responses in both languages demonstrate a reasonable, but not clear understanding of the concepts being asked in the prompts.
Science
Haitian-Creole
Item: Tadpole/Frog
All animals need oxygen to live.
Describe how a tadpole gets oxygen into its body.

**English Response:**
Maybe it breathe but you can't see that it breathe and maybe it breathe in it's body.

**Translation from Native Language:**
It takes oxygen in its mouth because it lives in water.

Describe how a frog gets oxygen into its body.

**English Response:**
maybe it's lungs grab oxegen but not through it's nose. But from another hole.

**Translation from Native Language:**
It takes oxygen in its nose because it lives on earth.
**Item Scores***:

<table>
<thead>
<tr>
<th>English Response (Tadpole, Frog)</th>
<th>Native Language Response (Tadpole, Frog)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scorer 1: 1, 2</td>
<td>Scorer 1: 1, 1</td>
</tr>
<tr>
<td>Scorer 2: 1, 2</td>
<td>Scorer 2: 2, 1</td>
</tr>
<tr>
<td>Mean Score: 1, 2</td>
<td>Mean Score: 1.5, 1</td>
</tr>
</tbody>
</table>

* *Item Scores*: 0 Minimum to 3 Maximum

**Commentary:**

In the English version, the student identifies lungs as the breathing structure of a frog. Because of its sound in Haitian-Creole, "e" is used in place of "y" in oxygen.

In the Haitian-Creole version, there is a slight indication that the student identifies the source of oxygen.

There is no strong evidence in either language that the student understands the concepts, or that she knows the amphibian’s respiratory structures and functions. The major mean score differences across languages is for Frog (2 and 1, respectively, in English and Haitian-Creole) due to the fact that knowledge of the frog’s breathing organs is not demonstrated consistently across languages.
### Teacher Ratings*

<table>
<thead>
<tr>
<th>Fluency in 1st Language</th>
<th>Literacy in 1st Language</th>
<th>Fluency in 2nd Language</th>
<th>Literacy in 2nd Language</th>
<th>Science Proficiency</th>
<th>Math Proficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>3</td>
<td>4</td>
<td>*</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

1. All animals need oxygen to live.

Describe how a tadpole gets oxygen into its body.

**English Response:**

It breath trough it nose in order to get oxygen and in there mouth.

**Translation from Native Language:**

It breathe through it’s mouth.

Describe how a frog gets oxygen into its body.

**English Response:**

The frog get it oxygen from open his/her mouth.

**Translation from Native Language:**

It breathe though it’s nose.
### Item Scores:

<table>
<thead>
<tr>
<th>Item</th>
<th>English Response (Tadpole, Frog)</th>
<th>Native Language Response (Tadpole, Frog)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scorer 1</td>
<td>1, 1</td>
<td>1, 1</td>
</tr>
<tr>
<td>Scorer 2</td>
<td>1, 1</td>
<td>1, 1</td>
</tr>
<tr>
<td>Mean Score</td>
<td>1, 1</td>
<td>1, 1</td>
</tr>
</tbody>
</table>

*Item Scores: 0 Minimum to 3 Maximum*

### Commentary:

The student uses Creole syntax to formulate the sentences in English. In addition to making spelling mistakes, the student code-switches frequently. The responses are more succinct and precise in Haitian-Creole than English.

Although the student clearly identifies two different ways of obtaining oxygen, in both languages she fails to identify gills and lungs, respectively, as the breathing structures of tadpoles and frogs. She also fails to describe correctly the source of oxygen.

The student does not have a complete understanding of the concepts in either language, which is consistent with the fact that in both languages the student receives the same low scores.
1. All animals need oxygen to live. Describe how a tadpole gets oxygen into its body.

**English Response:**

*It's because tadpole don't need any air to sink.*

**Translation from Native Language:**

*I think that it puts it in its mouth.*

Describe how a frog gets oxygen into its body.

**English Response:**

*No response given.*

**Translation from Native Language:**

*When it does "rebe" "rebe"*
**Item Scores**: 0 Minimum to 3 Maximum

<table>
<thead>
<tr>
<th>English Response (Tadpole, Frog)</th>
<th>Native Language Response (Tadpole, Frog)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scorer 1: 1, 0</td>
<td>Scorer 1: 1, 1</td>
</tr>
<tr>
<td>Scorer 2: 1, 0</td>
<td>Scorer 2: 1, 1</td>
</tr>
<tr>
<td>Mean Score: 1, 0</td>
<td>Mean Score: 1, 1</td>
</tr>
</tbody>
</table>

The student depends heavily on code-switching to provide his responses. There is a strong influence of the English language in the Haitian-Creole response. For example, “bout” in the Haitian-Creole response results from combining “mouth” and “bouch.”

“Rebe, rebe” in the response in Haitian-Creole indicates the sound frogs make.

The responses in both languages are inaccurate. However, the English response is poorer than the Haitian-Creole response.

Although the description of the breathing structure is more precise in the student’s native language, the source of oxygen and the organs used for breathing in tadpoles and frogs are not properly identified.
All animals need oxygen to live. Describe how a tadpole gets oxygen into its body.

**English Response:**

Tadpole. This is the animals. She feed. She mach. She breath

**Translation from Native Language:**

The animals eats. The animal drinks. The animal breathes. The animal walks. That is how the tadpole gets oxygen.

Describe how a frog gets oxygen into its body.

**English Response:**

No response given.

**Translation from Native Language:**

The frog lives under water.
<table>
<thead>
<tr>
<th>Item Scores*</th>
<th>Notes:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>English Response</strong></td>
<td><strong>Native Language Response</strong></td>
</tr>
<tr>
<td><em>(Tadpole, Frog)</em></td>
<td><em>(Tadpole, Frog)</em></td>
</tr>
<tr>
<td>Scorer 1: 1, 0</td>
<td>Scorer 1: 1, 1</td>
</tr>
<tr>
<td>Scorer 2: 1, 0</td>
<td>Scorer 2: 1, 2</td>
</tr>
<tr>
<td>Mean Score: 1, 0</td>
<td>Mean Score: 1, 1.5</td>
</tr>
</tbody>
</table>

* Item Scores: 0 Minimum to 3 Maximum

**Commentary:**

Although the Haitan-Creole response is more extensive than the English version, the response on both language versions is inaccurate. Native language words are used to substitute for English. For example “mache” is used in lieu of “walks.”
All animals need oxygen to live. Describe how a tadpole gets oxygen into its body.

**English Response:**
They getting from sun and it breathe.

**Translation from Native Language:**
They breathe under water, they have something in the rocks which helps them breathe. Their lungs help them breathe.

Describe how a frog gets oxygen into its body.

**English Response:**
They getting from the water and sun.

**Translation from Native Language:**
They have energy, they have lungs to breathe from. Their mother give them oxygen and energy and this way they are able to breathe.
### Item Scores:

<table>
<thead>
<tr>
<th>English Response</th>
<th>Native Language Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Tadpole, Frog)</td>
<td>(Tadpole, Frog)</td>
</tr>
<tr>
<td>Scorer 1: 1, 1</td>
<td>Scorer 1: 1, 2</td>
</tr>
<tr>
<td>Scorer 2: 2, 1</td>
<td>Scorer 2: 2, 2</td>
</tr>
<tr>
<td>Mean Score: 2, 1.5</td>
<td>Mean Score: 1.5, 2</td>
</tr>
</tbody>
</table>

* Item Scores: 0 Minimum to 3 Maximum

### Commentary:

The English influence in the student's native language is evident in the words “roc” ('rock') and ‘lung’ found embedded within the Haitian-Creole response. The breathing structure of tadpoles is not identified properly and there are misconceptions about the source of oxygen in either language.

The response in the student's language is richer. She demonstrates her understanding of respiration in frogs and tadpoles better in Haitian-Creole than English. However, the scores obtained in both languages are comparable.
Science
Spanish
Item: Erosion
The pictures below show the same river and mountains, but one picture shows how they looked millions of years ago, and the other picture shows how they look now. Circle the letter under the picture that shows how they look now. Explain how you can tell this.

<table>
<thead>
<tr>
<th>English Response:</th>
<th>Translation from Native Language:</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Picture B]. I could tell is that one is how the mountains are today.</td>
<td>[Picture B]. Because the mountains looked different before from now.</td>
</tr>
</tbody>
</table>

* Rating Scale
1 Minimum
5 Maximum
**Item Scores**: 0 Minimum to 3 Maximum

<table>
<thead>
<tr>
<th>English Response</th>
<th>Native Language Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scorer 1: 1</td>
<td>Scorer 1: 1</td>
</tr>
<tr>
<td>Scorer 2: 1</td>
<td>Scorer 2: 1</td>
</tr>
<tr>
<td>Mean Score: 1</td>
<td>Mean Score: 1</td>
</tr>
</tbody>
</table>

**Commentary:**

Both the Spanish and English responses are erroneous and very similar in content. The student selects the wrong picture and limits himself to restating the prompt into the response.

Spelling and sentence structure in Spanish is limited. For example, the student uses “montanias” instead of “montanas” replacing letters to write phonetically the word with English spelling influence. Also, “beian” should be “veian”.

The English response also identifies the wrong answer and shows limited grammar and spelling by the use of “be” instead of “because” and incomplete punctuation.
2. The pictures below show the same river and mountains, but one picture shows how they looked millions of years ago, and the other picture shows how they look now. Circle the letter under the picture that shows how they look now. Explain how you can tell this.

**English Response:**

[Picture B].
Because theirs more mountains and it makes more water.

**Translation from Native Language:**

[Picture B].
Because the river of today has more mountains and it makes more water.
Item Scores:

<table>
<thead>
<tr>
<th>English Response</th>
<th>Native Language Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scorer 1: 1</td>
<td>Scorer 1: 1</td>
</tr>
<tr>
<td>Scorer 2: 1</td>
<td>Scorer 2: 1</td>
</tr>
<tr>
<td>Mean Score: 1</td>
<td>Mean Score: 1</td>
</tr>
</tbody>
</table>

* Item Scores: 0 Minimum to 3 Maximum

Commentary:

Both the English and Spanish responses demonstrate very limited understanding of erosion. The student fails to identify the proper response and to give a reasonable explanation based on that concept.

There are grammatical and syntax errors in both languages. For example, in the Spanish version, he spells “aora” for “ahora” and “ase” for “hace”; and in English he spells “there” as “theirs.”

The responses are very similar in both languages, with no elaborate or clear understanding.
2. The pictures below show the same river and mountains, but one picture shows how they looked millions of years ago, and the other picture shows how they look now. Circle the letter under the picture that shows how they look now. Explain how you can tell this.

[Picture A]. Because in letter B the mountains could break down and in letter A the mountains can't change once its smood it can't be pointe.
<table>
<thead>
<tr>
<th>Item Scores*</th>
<th>Mean Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>English Response</td>
<td>Native Language Response</td>
</tr>
<tr>
<td>Scorer 1: 1</td>
<td>Scorer 1: 1</td>
</tr>
<tr>
<td>Scorer 2: 2</td>
<td>Scorer 2: 1</td>
</tr>
<tr>
<td>Mean Score: 1.5</td>
<td>Mean Score: 1</td>
</tr>
</tbody>
</table>

* Item Scores: 0 Minimum to 3 Maximum

Commentary:

Although poor, the English response is more detailed than the response in Spanish. In English, erosion is identified by the use of vocabulary substitution, through the use of "break-down" as a definition of eroding material. The comparison to the mountains on A, described as "smood" ("smooth" spelled by Spanish sound influence), as well as the conclusion that mountains cannot be "pointe" (pointy) after having been eroded, gives insight into the student's understanding of the concepts in the task.

Sentence structure and spelling in both languages are still in development. For example, in Spanish there is a lack of proper punctuation and poor spelling ("combertir" should be "convertir").
Estela

The pictures below show the same river and mountains, but one picture shows how they looked millions of years ago, and the other picture shows how they look now. Circle the letter under the picture that shows how they look now. Explain how you can tell this.

**English Response:**

[Picture B]

I know this because now the mountains are pointy and letter B is pointy, but letter A is like hills.

**Translation from Native Language:**

[Picture B]

I know this because now the mountains are pointy and letter B has points, but letter A does not have points. [student inserts a graphic of pointy mountains in the sentence]
Item Scores:

<table>
<thead>
<tr>
<th>English Response</th>
<th>Native Language Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scorer 1:  1</td>
<td>Scorer 1:  1</td>
</tr>
<tr>
<td>Scorer 2:  1</td>
<td>Scorer 2:  1</td>
</tr>
<tr>
<td>Mean Score:  1</td>
<td>Mean Score:  1</td>
</tr>
</tbody>
</table>

* Item Scores: 0 Minimum to 3 Maximum

Commentary:

The student provides erroneous and similar responses in both languages. The comparisons between options A and B do not provide adequate reasoning nor the proper answer. The student does not demonstrate an understanding of the concept of erosion in either language.

The English response is more coherent, using vocabulary and structure more fluently. The Spanish version shows immature sentence structure, limited vocabulary, and imprecise words, such as “puntos” (“points”), instead of “puntas” (“peaks”).
Science

Chinese

Item: Erosion
2. The pictures below show the same river and mountains, but one picture shows how they looked millions of years ago, and the other picture shows how they look now. Circle the letter under the picture that shows how the river and mountains look now. Explain how you can tell this.

**Picture A.**

The mountains wouldn't be so deep.

**Translation from Native Language:**

[Picture A]. It's because the role of erosion from long time ago to the present. Because it began many years ago, therefore today we have erosion.
**Item Scores**: 0 Minimum to 3 Maximum

<table>
<thead>
<tr>
<th>English Response</th>
<th>Native Language Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scorer 1: 2</td>
<td>Scorer 1: 3</td>
</tr>
<tr>
<td>Scorer 2: 2</td>
<td>Scorer 2: 3</td>
</tr>
<tr>
<td>Mean Score: 2</td>
<td>Mean Score: 3</td>
</tr>
</tbody>
</table>

**Commentary:**

The response in English demonstrates understanding of the concept addressed, showing limited vocabulary development and a working command of the language.

Even though the Chinese sentence structure is still developing and the vocabulary is at a word recognition level, the student's response in Chinese is more complete than the English version. This can be seen through the student's use of “erosion” in context as part of the answer. This is reflected in a higher score for the Chinese response.
### Teacher Ratings:

<table>
<thead>
<tr>
<th>Fluency in 1st Language</th>
<th>Literacy in 1st Language</th>
<th>Fluency in 2nd Language</th>
<th>Literacy in 2nd Language</th>
<th>Science Proficiency</th>
<th>Math Proficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

2. The pictures below show the same river and mountains, but one picture shows how they looked millions of years ago, and the other picture shows how they look now. Circle the letter under the picture that shows how they look now. Explain how you can tell this.

**English Response:**

Millions of years ago of river A, there mouth are very low and clean. The mouths like people, they can grow big each year. They grow very tall after Millions year ago. Therefore, I choose B.

**Translation from Native Language:**

[Picture B]. After millions of years, little hills grew into big mountains. Therefore, I choose B.
**Item Scores**: 0 Minimum to 3 Maximum

<table>
<thead>
<tr>
<th>English Response</th>
<th>Native Language Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scorer 1: 1</td>
<td>Scorer 1: 1</td>
</tr>
<tr>
<td>Scorer 2: 1</td>
<td>Scorer 2: 1</td>
</tr>
<tr>
<td>Mean Score: 1</td>
<td>Mean Score: 1</td>
</tr>
</tbody>
</table>

**Commentary:**

The English response shows vocabulary and language development, with a circular form of writing, repetitive structure, and emerging syntax.

The Chinese response is clearer and better-structured than the English version.

However, the response is erroneous in both languages and shows misconceptions (e.g., “mountains grow like people”). The student has no conceptual understanding of erosion or the processes that change landforms over time. The response is scored continually low across both languages.
The pictures below show the same river and mountains, but one picture shows how they looked millions of years ago, and the other picture shows how they look now. Circle the letter under the picture that shows how they look now. Explain how you can tell this.

English Response:

[Picture A].
because it has little erosion and the other had bigger erosion

Translation from Native Language:

[Picture A].
Because the river eroded very small pieces of greenland.
**Item Scores**:  

<table>
<thead>
<tr>
<th>English Response</th>
<th>Native Language Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scorer 1: 2</td>
<td>Scorer 1: 3</td>
</tr>
<tr>
<td>Scorer 2: 3</td>
<td>Scorer 2: 3</td>
</tr>
<tr>
<td>Mean Score: 2.5</td>
<td>Mean Score: 3</td>
</tr>
</tbody>
</table>

*Item Scores: 0 Minimum to 3 Maximum*

**Commentary:**

There is less specificity in the English version. Even though the word erosion is used, limited sentence structure and emerging vocabulary prevents the student from conveying a clear comparison between the diagrams, leaving the reader to ponder whether or not the concept is fully understood.

The Chinese response includes a clearer description of erosion than the comparison established in the English response. It is evident that the student is able to communicate the concept being addressed more eloquently in Chinese than in English, by giving a simple definition of erosion as taking tiny pieces of land away.
Science
Haitian-Creole
Item: Erosion
2. The pictures below show the same river and mountains, but one picture shows how they looked millions of years ago, and the other picture shows how they look now. Circle the letter under the picture that shows how the river and mountains look now. Explain how you can tell this.

**English Response:**

I pick A Because mountains are not cacky and how mountains are open to let the river go through.

**Translation from Native Language:**

[Picture A]. I choose B because the mountains do not look smooth.
**Item Scores**: 0 Minimum to 3 Maximum

<table>
<thead>
<tr>
<th>English Response</th>
<th>Native Language Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scorer 1:</td>
<td>Scorer 1:</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Scorer 2:</td>
<td>Scorer 2:</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Mean Score:</td>
<td>Mean Score:</td>
</tr>
<tr>
<td>2.5</td>
<td>1.5</td>
</tr>
</tbody>
</table>

Commentary:

The student does not show proper understanding of the concepts addressed in either language. However the combinations of strengths and weaknesses are different in each language.

In the English version, the student identifies the correct picture, but her use of language is imprecise and the justification of her selection is incorrect. By contrast, in the Haitian-Creole version, the student identifies the wrong picture, but her justification is a bit clearer.

Probably as a result of these different combinations of weaknesses and strengths, a total consistency of scoring across raters is not accomplished and the mean scores across languages are different.
Science

Spanish

Item: Heat
3. A nail becomes warm when it is hammered into a piece of wood. Tell why the nail becomes warm.

translation from native language:

A nail becomes warm because there is friction in the wood that makes it heat up.

English response:

Epublication per clavo se calienta cuando se martilla en una tabla de madera.

3. Un clavo se pone caliente cuando es martillado en una tabla de madera. Explique por qué el clavo se calienta.

Translation from native language:

A nail becomes warm when it is hammered into a piece of wood because there is friction in the wood.

English response:

A nail becomes warm when it is hammered into a piece of wood because of the friction in the wood.
**Item Scores**:  

<table>
<thead>
<tr>
<th>English Response</th>
<th>Native Language Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scorer 1:</td>
<td>Scorer 1:</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Scorer 2:</td>
<td>Scorer 2:</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Mean Score:</td>
<td>Mean Score:</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

* Item Scores: 0 Minimum to 3 Maximum

**Commentary:**

In both the English and Spanish versions, the student demonstrates some understanding of the transfer of energy processes taking place, by mentioning friction as a factor. However, the student's explanation in both languages implies that friction is in the wood.

It could be debated in a literal translation that the action of resistance between the wood fibers and the nail is happening within the wood. Nonetheless, the responses lack more explanation and detail that would show conceptual understanding. The student seems equally fluent in both languages. He transposes the English “friction” into Spanish.

Although the qualities of the responses in both languages are comparable, the scorers scored the response in Spanish lower than the response in English.
A nail becomes warm when it is hammered into a piece of wood. Tell why the nail becomes warm.

**English Response:**
Because the wood when the sun shines in it on a hot day it gets hot and the nail it's hamered it to the piece of wood it gets warm.

**Translation from Native Language:**
Because when the sun warms the wood and when it is hammered it gets hot.
**Item Scores**:  

<table>
<thead>
<tr>
<th></th>
<th>English Response</th>
<th>Native Language Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scorer 1:</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Scorer 2:</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Mean Score:</td>
<td>1.5</td>
<td>1.5</td>
</tr>
</tbody>
</table>

* Item Scores: 0 Minimum to 3 Maximum

**Commentary:**

The responses in both languages reflect different combinations of weaknesses.

For example, in the English version the response has syntax and grammatical errors (e.g., "Because the wood when the sun shines..."); and in Spanish "una" ("fingernail") is used improperly to refer to a nail.

In both languages, the responses reflect the misconception that the nail becomes warm because the wood is warm. There is the assumption that the nails are being hammered under the sun, which causes them to heat up. The responses in both languages are scored low.
### Teacher Ratings*

<table>
<thead>
<tr>
<th>Fluency in 1st Language</th>
<th>Literacy in 1st Language</th>
<th>Fluency in 2nd Language</th>
<th>Literacy in 2nd Language</th>
<th>Science Proficiency</th>
<th>Math Proficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

3. Un clavo se pone caliente cuando es martillado en una tabla de madera.  
Explica por qué el clavo se calienta.

**English Response:**  
because you hit it hard ashes mite come out like bite of fire and that get it warm.

**Translation from Native Language:**  
The nail gets hot because when they hit it hard sparks can jump out and this makes the nail hot.

---

### Carlos

A nail becomes warm when it is hammered into a piece of wood.  
Tell why the nail becomes warm.

**English Response:**  
because you hit it hard ashes mite come out like bite of fire and that get it warm.

---

* Rating Scale  
1 Minimum to  
5 Maximum
Item Scores*:

<table>
<thead>
<tr>
<th>English Response</th>
<th>Native Language Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scorer 1: 2</td>
<td>Scorer 1: 1</td>
</tr>
<tr>
<td>Scorer 2: 1</td>
<td>Scorer 2: 2</td>
</tr>
<tr>
<td>Mean Score: 1.5</td>
<td>Mean Score: 1.5</td>
</tr>
</tbody>
</table>

* Item Scores: 0 Minimum to 3 Maximum

Commentary:

The responses demonstrate a limited language foundation. There are spelling, grammatical, and syntax errors in both languages. (e.g., use of "quando" for "cuando"; "come out like a bite of fire"). Notice the use of an imprecise word, "ashes," to refer to sparks in the English version.

The explanations given in both language versions are similar—and incorrect. The student does not show understanding of the concept of heat in either language.
A nail becomes warm when it is hammered into a piece of wood.
Tell why the nail becomes warm.

English Response: The nails become warm because if you hit it in to a peace of wood when it goes all the way down its going to become warm.

Translation from Native Language: The nail gets warmer when somebody hammers it on a piece of wood all the way in.
**Item Scores**: 0 Minimum to 3 Maximum

<table>
<thead>
<tr>
<th>English Response</th>
<th>Native Language Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scorer 1: 3</td>
<td>Scorer 1: 1</td>
</tr>
<tr>
<td>Scorer 2: 1</td>
<td>Scorer 2: 1</td>
</tr>
<tr>
<td>Mean Score: 2</td>
<td>Mean Score: 1</td>
</tr>
</tbody>
</table>

**Commentary:**

The English and the Spanish responses show limited understanding of the concept of heat. In both languages, the student only restates part of the prompt, while emphasizing the idea that the depth of the nail hammered into the wood is the factor which causes the nail to heat up.

The high score given by scorer 1 to the English response is definitely due to a misinterpretation of the scoring rubrics or the student's response.
Science
Chinese
Item: Heat
### Teacher Ratings:

<table>
<thead>
<tr>
<th>Fluency in 1st Language</th>
<th>4</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Literacy in 1st Language</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Fluency in 2nd Language</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Literacy in 2nd Language</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Science Proficiency</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Math Proficiency</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

### A nail becomes warm when it is hammered into a piece of wood. 
Tell why the nail becomes warm.

**English Response:**

3. A nail becomes warm when it is hammered into a piece of wood.

Tell why the nail becomes warm.

Maybe the wood is so hard. The nail can't put in. Then it hammered and it became warm.

**Translation from Native Language:**

Because of resistance and friction the nail becomes warm.
**Item Scores**:

<table>
<thead>
<tr>
<th>English Response</th>
<th>Native Language Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scorer 1: 1</td>
<td>Scorer 1: 2</td>
</tr>
<tr>
<td>Scorer 2: 1</td>
<td>Scorer 2: 2</td>
</tr>
<tr>
<td>Mean Score: 1</td>
<td>Mean Score: 2</td>
</tr>
</tbody>
</table>

* Item Scores: 0 Minimum to 3 Maximum

**Commentary:**

The English version shows limited understanding of the concepts being addressed, limited vocabulary to convey meaning, and emerging sentence structure.

In the Chinese version, the student does not mention energy transfer but includes resistance and friction as part of the explanation.

The student's response in Chinese is more accurate and complete. This is reflected in a higher score for the Chinese response.
Kang Han

A nail becomes warm when it is hammered into a piece of wood.
Tell why the nail becomes warm.

English Response:
When a nail hammered into a piece of wood, you have to hit very hard. When hammer touch nail very hard, nail will be hot

Translation from Native Language:
Friction makes objects warm. When a nail is hammered into a piece of wood, the nail was acted upon by strong friction, and therefore, it gets warm.
Item Scores*:

<table>
<thead>
<tr>
<th>English Response</th>
<th>Native Language Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scorer 1: 1</td>
<td>Scorer 1: 3</td>
</tr>
<tr>
<td>Scorer 2: 2</td>
<td>Scorer 2: 2</td>
</tr>
<tr>
<td>Mean Score: 1.5</td>
<td>Mean Score: 2.5</td>
</tr>
</tbody>
</table>

* Item Scores: 0 Minimum to 3 Maximum

Commentary:

The English response shows emerging sentence structure and development of vocabulary in context. The sentence structure and Chinese characters used show command of the language. In both languages, the student fails to answer in terms of energy transformations. The concept of friction does not appear in the English response, minimizing the level of understanding the student is showing in his primary language.

The response in Chinese is more articulate than in English and shows more clear understanding of the concepts (friction heat caused by friction).
A nail becomes warm when it is hammered into a piece of wood. Tell why the nail becomes warm.

I think a nail becomes warm because a wood have warm in it. So when the nail went in the wood get warm to the nail, so the nail become warm.

Because the hammer hit the nail on the wood, there was heat inside the wood. So the nail would be heated.

Ming Sui

3. A nail becomes warm when it is hammered into a piece of wood. Tell why the nail becomes warm.

I think a nail becomes warm because a wood have warm in it. So when the nail went in the wood get warm to the nail, so the nail become warm.

Because the hammer hit the nail on the wood, there was heat inside the wood. So the nail would be heated.
### Item Scores*

<table>
<thead>
<tr>
<th>English Response</th>
<th>Native Language Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scorer 1: 1</td>
<td>Scorer 1: 1</td>
</tr>
<tr>
<td>Scorer 2: 1</td>
<td>Scorer 2: 1</td>
</tr>
<tr>
<td>Mean Score: 1</td>
<td>Mean Score: 1</td>
</tr>
</tbody>
</table>

* Item Scores: 0 Minimum to 3 Maximum

### Commentary:

The student's responses in both languages are comparable and incorrect. There is no evidence of understanding the concepts addressed in the prompt. The English version shows circular writing, repetitive language use, and emerging sentence structure and spelling.
Science
Haitian-Creole
Item: Heat
A nail becomes warm when it is hammered into a piece of wood. Tell why the nail becomes warm.

**English Response:**

Because when it is nailed in the wood make a hole and whehe it makes a hole it come kind of hot and when you nailed the nailes in it becomes warm.

**Translation from Native Language:**

Because the nail is metal when you hit the metal into the wood, the metal becomes hot.
**Item Scores**: 0 Minimum to 3 Maximum

<table>
<thead>
<tr>
<th>English Response</th>
<th>Native Language Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scorer 1: 2</td>
<td>Scorer 1: 2</td>
</tr>
<tr>
<td>Scorer 2: 2</td>
<td>Scorer 2: 3</td>
</tr>
<tr>
<td>Mean Score: 2</td>
<td>Mean Score: 2.5</td>
</tr>
</tbody>
</table>

**Commentary:**

The sentence structure in the English response shows the influence of Haitian-Creole syntax.

The notion of friction is not present in the response on either version, but there is some understanding that heat is produced as a form of energy.

Although the responses in both languages are comparable, the response to the Haitian-Creole version is more concise and articulate.

Inconsistency in the scoring across languages is due to the fact that scorer 2 scored the Haitian-Creole response a little too high.
A nail becomes warm when it is hammered into a piece of wood. Tell why the nail becomes warm.

Charles

Because it metal and went metal touch it tured hot

because it is a metal
### Item Scores*

<table>
<thead>
<tr>
<th>English Response</th>
<th>Native Language Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scorer 1: 1</td>
<td>Scorer 1: 1</td>
</tr>
<tr>
<td>Scorer 2: 1</td>
<td>Scorer 2: 1</td>
</tr>
<tr>
<td>Mean Score: 1</td>
<td>Mean Score: 1</td>
</tr>
</tbody>
</table>

* Item Scores: 0 Minimum to 3 Maximum

---

**Commentary:**

The student uses Haitian-Creole syntax to formulate the English sentence response. Words such as “beski” in lieu of “because” is a combination of English and Haitian language (“peski”).

The responses on both language versions are insufficient and demonstrate limited understanding of the concept.

The student attempts a more elaborate explanation in English, although the spelling and the syntax in the language is very limited.
Mathematics

Spanish

Item: Figures
3. A fourth-grade class needs 5 leaves each day to feed its 5 caterpillars. How many leaves would they need each day for 12 caterpillars?

Answer: Use drawings, words, or numbers to show how you got your answer.

Think carefully about the following question. Write a complete answer. You may use drawings, words, and numbers to explain your answer. Be sure to show all of your work.

1. In what ways are the figures above alike? List as many ways as you can.

because their the same number and their are different shapes and they got the number

In what ways are the figures above different? List as many ways as you can.

one of the shapes are different then the other shap

* Rating Scale
1  Minimum
5  Maximum
### Item Scores*:

<table>
<thead>
<tr>
<th>English Response</th>
<th>Native Language Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scorer 1: 3</td>
<td>Scorer 1: 3</td>
</tr>
<tr>
<td>Scorer 2: 2</td>
<td>Scorer 2: 3</td>
</tr>
<tr>
<td>Mean Score: 2.5</td>
<td>Mean Score: 3</td>
</tr>
</tbody>
</table>

* Item Scores: 0 Minimum to 5 Maximum

### Commentary:

Emerging syntax and vocabulary are evident in the use of "diff't" for "different," "nomeber" for "number," and "shap" for "shape." The student writes "dedos" ("fingers") instead of "lados" ("sides").

Although the Mayan roots of this student may influence his use of Spanish, he uses more detailed mathematical language in Spanish. Thus, the Spanish response is scored higher and more consistently than the English response.
Teacher Ratings:

<table>
<thead>
<tr>
<th></th>
<th>Fluency in 1st Language</th>
<th>Literacy in 1st Language</th>
<th>Fluency in 2nd Language</th>
<th>Literacy in 2nd Language</th>
<th>Science Proficiency</th>
<th>Math Proficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Language</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

Marcos

In what ways are the figures above and below alike?
List as many ways as you can.

English Response:

Translation from Native Language:

1. They have four sides
2. They are polygons
3. They have edges
4. They have vertices

In what ways are the figures above and below different?
List as many ways as you can.

English Response:

Translation from Native Language:

1. One is slanted, the other is not.
2. One is a rectangle, the other is not.
**Item Scores:**

<table>
<thead>
<tr>
<th>English Response</th>
<th>Native Language Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scorer 1: 2</td>
<td>Scorer 1: 3</td>
</tr>
<tr>
<td>Scorer 2: 2</td>
<td>Scorer 2: 3</td>
</tr>
<tr>
<td>Mean Score: 2</td>
<td>Mean Score: 3</td>
</tr>
</tbody>
</table>

* Item Scores: 0 Minimum to 5 Maximum

---

**Commentary:**

Although the student appears to have several terms in the English language relating to geometric figures, they are misused in parts of the answer. The student can provide a clearer and more accurate response in Spanish. Terms used in the English version are not transferred to the Spanish language response (i.e.: "vertices," "polígonos").

The student demonstrates knowledge of quadrilaterals' attributes by mentioning corners. By describing one shape as a rectangle, and the other as "not," he shows an understanding for definition. He incorrectly calls the figure a trapezoid, while inconsistently using descriptions, such as naming the figures "four-sided" and then "one-sided".

In the Spanish version, he accurately compares and contrasts at least two attributes for each figure, using numbers to make a list and responding in full sentences.

The scoring is consistent with these observations. The Spanish version was scored higher than the English response.
Think carefully about the following question. Write a complete answer. You may use drawings, words, and numbers to explain your answer. Be sure to show all of your work.

1. In what ways are the figures above alike? List as many ways as you can.

In what ways are the figures above different? List as many ways as you can.

Translation from Native Language:

One is vertically and one is orysandy. One is stait and one is a side.

Because one is leaning and the other is straight.
<table>
<thead>
<tr>
<th>Item Scores*:</th>
<th>Notes:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Item Scores</strong>: 0 Minimum to 5 Maximum</td>
<td></td>
</tr>
<tr>
<td><strong>English Response</strong></td>
<td><strong>Native Language Response</strong></td>
</tr>
<tr>
<td>Scorer 1: 4</td>
<td>Scorer 1: 2</td>
</tr>
<tr>
<td>Scorer 2: 4</td>
<td>Scorer 2: 1</td>
</tr>
<tr>
<td>Mean Score: 4</td>
<td>Mean Score: 1.5</td>
</tr>
</tbody>
</table>

**Commentary:**

Student responses in both the Spanish and English versions are limited by the use of incorrect terms in both languages. The student uses positional-space relationship descriptors as the main variable. However, shape and height are included in the English version, making this response more accurate. Thus, the English response scored consistently higher.
Think carefully about the following question. Write a complete answer. You may use drawings, words, and numbers to explain your answer. Be sure to show all of your work.

1. In what ways are the figures above and below alike? List as many ways as you can.

   English Response:

   There are different in many ways like square and a hexagon. The first one is a square and the other one is a hexagon.

   Translation from Native Language:

   The first one is a crooked rectangle and the other one is stretched.

2. In what ways are the figures above and below different? List as many ways as you can.

   English Response:

   The first square is strait and the other one is a hexagon.

   Translation from Native Language:

   The first is deformed and the other one is straight.
The student shows limited and erroneous use of English terms to differentiate the figures. The Spanish response, although limited, is more accurate. For example the student does not identify a shape erroneously as a hexagon. There is a basic understanding of the concepts being asked in both languages. Notwithstanding, the raters gave equal scores across languages and consistently varied in scores within each language.
Mathematics

Chinese

Item: Figures
Think carefully about the following question. Write a complete answer. You may use drawings, words, and numbers to explain your answer. Be sure to show all of your work.

In what ways are the figures above alike? List as many ways as you can.

**English Response:**
- Their both a rectangle.
- They both have 18 squares inside the rectangle.
- They both are 3 inches.
- They both could put 4 triangles inside.

**Translation from Native Language:**
- There are 4 triangles in both diagrams.
- Both of them have straight lines.
- Both of have 360°.
- Both of them have 18 small square.

In what ways are the figures above different? List as many ways as you can.

**English Response:**
- One is flat down.
- One triangle inside is straight down.
- Some of the small squares inside are half.

**Translation from Native Language:**
- One of them is slanted.
- One of them is flat.
- One of them is a parallelogram. They are different in shape.

---

*Rating Scale*

1. Minimum
2. Minimum
3. Minimum
4. Minimum
5. Maximum
<table>
<thead>
<tr>
<th>Item Scores*</th>
<th>Notes:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>English Response</strong></td>
<td><strong>Native Language Response</strong></td>
</tr>
<tr>
<td>Scorer 1: 4</td>
<td>Scorer 1: 5</td>
</tr>
<tr>
<td>Scorer 2: 4</td>
<td>Scorer 2: 4</td>
</tr>
<tr>
<td>Mean Score: 4</td>
<td>Mean Score: 4.5</td>
</tr>
</tbody>
</table>

* Item Scores: 0 Minimum to 5 Maximum

**Commentary:**

Although the student's response in Chinese is more detailed and demonstrates more complex comparisons than the English version, the student still lacks some vocabulary and resorts to illustrations.

Both language responses show understanding of the concepts addressed. The English response uses descriptive language such as "flat down" instead of "diagonal" or "slanted" with the English being consistent across subject areas.
### Teacher Ratings*:

<table>
<thead>
<tr>
<th>Fluency in 1st Language</th>
<th>Literacy in 1st Language</th>
<th>Fluency in 2nd Language</th>
<th>Literacy in 2nd Language</th>
<th>Science Proficiency</th>
<th>Math Proficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

*Rating Scale
1 Minimum to 5 Maximum

Think carefully about the following question. Write a complete answer. You may use drawings, words, and numbers to explain your answer. Be sure to show all of your work.

1. **In what ways are the figures above alike?** List as many ways as you can.

   English Response:

   The area of them are same, lines are equal

   **Translation from Native Language:**

   Their length are the same. They are quadrilaterals. Their area and perimeter are the same. They are on the same base line. Their sides are parallel.

   In what ways are the figures above and below different? List as many ways as you can.

   English Response:

   The first one, all lines of it is straight like 1/1, if we went to know the area, it's not same. The name of them are different. Same way to know the area of them. Not placed together.

   **Translation from Native Language:**

   Their names are different. They placed differently. One of them is straight, the other is slanted. One of them his something in front and part at the back.
Item Scores:

<table>
<thead>
<tr>
<th>English Response</th>
<th>Native Language Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scorer 1: 3</td>
<td>Scorer 1: 3</td>
</tr>
<tr>
<td>Scorer 2: 5</td>
<td>Scorer 2: 5</td>
</tr>
<tr>
<td>Mean Score: 4</td>
<td>Mean Score: 4</td>
</tr>
</tbody>
</table>

* Item Scores: 0 Minimum to 5 Maximum

Commentary:

The Chinese description and comparisons are very detailed, clear, and correct. The student is able to use proper terms and gives four similarities and differences. Whereas in the English language, the student shows limited vocabulary use and thus a limited understanding of concepts addressed, resorting to embedded illustrations within the text conveys the meaning in the reasoning behind the response.

It is evident from the Chinese version that the student understands the concepts. The diagram representations complement the English entries, but they are not as comprehensive as the Chinese descriptions. Nevertheless, raters were consistent in scoring content criteria within the variations of both languages.
Mathematics
Haitian-Creole
Item: Figures
Think carefully about the following question. Write a complete answer. You may use drawings, words, and numbers to explain your answer. Be sure to show all of your work.

In what ways are the figures above and below alike?
List as many ways as you can.

English Response: Parallelogram
Translation from Native Language: They both have four sides.

In what ways are the figures above and below different?
List as many ways as you can.

English Response: One is crooked on is straight.
Translation from Native Language: One is oblique. One is straight.
**Item Scores**:  

<table>
<thead>
<tr>
<th>English Response</th>
<th>Native Language Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scorer 1: 4</td>
<td>Scorer 1: 3</td>
</tr>
<tr>
<td>Scorer 2: 3</td>
<td>Scorer 2: 3</td>
</tr>
<tr>
<td>Mean Score: 3.5</td>
<td>Mean Score: 3</td>
</tr>
</tbody>
</table>

* Item Scores: 0 Minimum to 5 Maximum

**Commentary:**

In the English version, the student identifies the name of the figure, but lacks the contextual vocabulary to identify its comparing features in detail.

The student gives one reason why the figures are alike and two reasons to explain how they are different, omitting final vowels in words and using word sounds for spelling.

There is very little elaboration of reasonings on the two language versions.

In the Haitian-Creole version, the student describes with adequate language two of the figures' contrasting characteristics and one similarity. These findings are reflected in the score of one of the raters, which is slightly higher than that of the other rater in the native language response, yet consistent in the English version.
Charles

In what ways are the figures above and below alike?
List as many ways as you can.

English Response: Both of them are quadrilateral. They have 2 parallel lines.

Translation from Native Language:
They have 2 parallel lines they are quadrilateral.

In what ways are the figures above and below different?
List as many ways as you can.

English Response: One have 4 right angles one has 2 obtuse angles.

Translation from Native Language:
One is bent one is straight.

Translation from Native Language:
Both of them are quadrilateral they have 2 pair of parallel lines.

Translation from Native Language:
One have 4 right angles one has 2 obtuse angles.
<table>
<thead>
<tr>
<th>Item Scores*:</th>
<th>Notes:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>English Response</strong></td>
<td><strong>Native Language Response</strong></td>
</tr>
<tr>
<td>Scorer 1:</td>
<td>4</td>
</tr>
<tr>
<td>Scorer 2:</td>
<td>5</td>
</tr>
<tr>
<td>Mean Score:</td>
<td>4.5</td>
</tr>
</tbody>
</table>

* Item Scores: 0 Minimum to 5 Maximum

**Commentary:**

The English responses are more complete, though syntax, grammar, and spelling show limited use. It is evident that the student uses the English language vocabulary to describe the figures in both testing situations. Nevertheless, the raters recognized the basic level of concept understandings, thus giving higher scores. One rater identified the richer mathematical language in the English version, giving it a considerably higher score.
Think carefully about the following question. Write a complete answer. You may use drawings, words, and numbers to explain your answer. Be sure to show all of your work.

In what ways are the figures above alike? List as many ways as you can.

**English Response:**
- They have 4 sides.
- They both have angles.

**Translation from Native Language:**
- They have 4 corners and 4 sides.
- They have parallel sides. The other doesn't. The rectangle has 4 right angles.

In what ways are the figures above different? List as many ways as you can.

**English Response:**
- I have a right angle. The other has a acute.

**Translation from Native Language:**
- They both have angles. The other has a right angle.
**Item Scores***:

<table>
<thead>
<tr>
<th>English Response</th>
<th>Native Language Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scorer 1: 5</td>
<td>Scorer 1: 3</td>
</tr>
<tr>
<td>Scorer 2: 4</td>
<td>Scorer 2: 3</td>
</tr>
<tr>
<td>Mean Score: 4.5</td>
<td>Mean Score: 3</td>
</tr>
</tbody>
</table>

* Item Scores: 0 Minimum to 5 Maximum

**Commentary:**

The student's response in English shows evidence of lack of punctuation and capitalization rules, as well as letter transposition; however, the student shows an understanding of the concept and attempts to utilize colloquial vocabulary to respond.

The student's Haitian-Creole response is more detailed than the English version, with a greater command of Haitian spelling and syntax.
Think carefully about the following question. Write a complete answer. You may use drawings, words, and numbers to explain your answer. Be sure to show all of your work.

1. In what ways are the figures above alike? List as many ways as you can.

English Response:

In above like line parallel.

Translation from Native Language:

I see a parallel and one line is slanted. They both have four sides with four angles.

2. In what ways are the figures above different? List as many ways as you can.

English Response:

One is parallel the other on is not strait parallel.

Translation from Native Language:

One is straight and the other is leaning to one side. One has four right angles and one has four slant- ed angles.
**Item Scores**: 0 Minimum to 5 Maximum

<table>
<thead>
<tr>
<th>English Response</th>
<th>Native Language Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scorer 1: 1</td>
<td>Scorer 1: 3</td>
</tr>
<tr>
<td>Scorer 2: 3</td>
<td>Scorer 2: 4</td>
</tr>
<tr>
<td>Mean Score: 2</td>
<td>Mean Score: 3.5</td>
</tr>
</tbody>
</table>

**Commentary:**

In the Haitian–Creole version, the student is able to provide more detailed descriptions than in the English version.

In the English version, some colloquial words replace mathematical terms such as “leaning” vs. “oblique” and “slanted angles” vs. “acute/obtuse angles.”

Also, the student uses conventions of spelling from Haitian-Creole, such as omitting the final silent “e” and substituting meanings of words from her native language to convey understanding.

Since Haitian children are seldom asked for their opinions and/or explanations, this type of open-ended item response is unfamiliar to them.

The response in English also shows limited concepts and fewer descriptions, as evidenced by the scores.

Within each language, scoring is inconsistent. However, the response in English is scored lower than the response in Haitian-Creole.
Mathematics
Spanish
**Item: Gumball Machine**
Think carefully about the following question. Write a complete answer. You may use drawings, words, and numbers to explain your answer. Be sure to show all of your work.

2. The gum ball machine has 100 gum balls; 20 are yellow, 30 are blue, and 50 are red. The gum balls are well mixed inside the machine.

Jenny gets 10 gum balls from the machine.

What is your best prediction of the number that will be red?

Answer: 5

Explain why you chose this number.

Because there is more balls that are red, that is why I think there is going to be 5 red balls.
The student's responses show a greater command of English than Spanish. In English, the student uses a math term "most," to say or explain that it is the highest number possible. There are still some grammatical errors, such as the omission of the verb "are" and the misplacement of "so," suggesting transitional skills and English language development.

Although the student was able to read the Spanish prompt, the spelling and grammar in the Spanish version show a weak foundation in her primary language. Syntax errors such as the use of "aiy" for "hay," "qureo" for "creo," and "ba" for "va," show limited Spanish writing skills. The responses are comparable across language versions.
Think carefully about the following question. Write a complete answer. You may use drawings, words, and numbers to explain your answer. Be sure to show all of your work.

The gum ball machine has 100 gum balls; 20 are yellow, 30 are blue, and 50 are red. The gum balls are well mixed inside the machine.

Jenny gets 10 gum balls from this machine.

What is your best prediction of the number that will be red?

Answer: 40 gum balls

Because des is 50 red, 30 blue, 20 yellow bubble gum balls take away 10 equal 90.

Because des is = reds
30 blue, 20 yellow bubble gum balls take away 10 equal 90.
### Item Scores*

<table>
<thead>
<tr>
<th>English Response</th>
<th>Native Language Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scorer 1: 2</td>
<td>Scorer 1: 2</td>
</tr>
<tr>
<td>Scorer 2: 1</td>
<td>Scorer 2: 1</td>
</tr>
<tr>
<td>Mean Score: 1.5</td>
<td>Mean Score: 1.5</td>
</tr>
</tbody>
</table>

* Item Scores: 0 Minimum to 5 Maximum

### Commentary:

Letter sounds from Spanish influence the spelling of English words (e.g., use of “des” for “this,” “buble” for “bubble”). Syntax is stronger in Spanish than in English.

The student uses a different problem-solution strategy in each language. In English, she totals all the gumballs and subtracts 10. In Spanish, she works only with the set of colored gumballs in the question, but still chooses to subtract 10 from the total of red gumballs. Her written response shows a total of 10 instead of the 40 marked on the answer line.

The responses in both languages are incorrect. The student shows very limited understanding of the concepts and operations needed to solve the problem. Although somewhat inconsistent within each language, the scores are low. The student obtains the same low scores in both languages.
Mathematics
Chinese
Item: Gumball Machine
Think carefully about the following question. Write a complete answer. You may use drawings, words, and numbers to explain your answer. Be sure to show all of your work.

2. The gum ball machine has 100 gum balls; 20 are yellow, 30 are blue, and 50 are red. The gum balls are well mixed inside the machine. Jenny gets 10 gum balls from this machine. What is your best prediction of the number that will be red?

Answer:

5.

If there were 30 blue candies in the machine, 50 in red and 20 in yellow, there were more red than blue and yellow ones. The chances of picking red was greater. So I think the answer was 5.
## Item Scores*

<table>
<thead>
<tr>
<th>English Response</th>
<th>Native Language Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scorer 1: 1</td>
<td>Scorer 1: 4</td>
</tr>
<tr>
<td>Scorer 2: 3</td>
<td>Scorer 2: 5</td>
</tr>
<tr>
<td>Mean Score: 2</td>
<td>Mean Score: 4.5</td>
</tr>
</tbody>
</table>

* Item Scores: 0 Minimum to 5 Maximum

## Commentary:

Although the response and reasoning are correct for both language versions, the reasonings are expressed with more detail in Chinese. In addition, the student does not specify the number of red gumballs in the blank space provided. As a result, the response is scored high in the student’s native language.
Think carefully about the following question. Write a complete answer. You may use drawings, words, and numbers to explain your answer. Be sure to show all of your work.

2. The gum ball machine has 100 gum balls; 20 are yellow, 30 are blue, and 50 are red. The gum balls are well mixed inside the machine. Jenny gets 10 gum balls from this machine.

What is your best prediction of the number that will be red?

Answer: 5

Because: 100 ÷ 10 = 10
If he get 10 out that mean 50 ÷ 10 = 5 and that is why I chose this number.
### Item Scores:

<table>
<thead>
<tr>
<th>English Response</th>
<th>Native Language Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scorer 1: 4</td>
<td>Scorer 1: 4</td>
</tr>
<tr>
<td>Scorer 2: 4</td>
<td>Scorer 2: 5</td>
</tr>
<tr>
<td>Mean Score: 4</td>
<td>Mean Score: 4.5</td>
</tr>
</tbody>
</table>

* Item Scores: 0 Minimum to 5 Maximum

### Commentary:

The student’s responses in English and Chinese have different sets of weaknesses. Whereas English sentence structure is still developing and Chinese influences the use of a third person singular conjugation (i.e., the “s” is omitted in “...if he get...”), the response in Chinese is more complex, although it lacks closure.

The student uses a different, but correct approach to solve the problem in each language. Whereas division is used in the English response, in the Chinese version the student uses a more complex explanation—(proportions) to justify his answer.

The student writes the proper number response in both languages; the ratings are almost identical for both languages.
The gum ball machine has 100 gum balls; 20 are yellow, 30 are blue, and 50 are red. The gum balls are well mixed inside the machine. Jenny gets 10 gum balls from the machine. What is your best prediction of the number that will be red?

**Translation from Native Language:**

五
She will maybe get five because there are more red then the rest of the color.

**English Response:**

five

Think carefully about the following question. Write a complete answer. You may use drawings, words, and numbers to explain your answer. Be sure to show all of your work.

2. The gum ball machine has 100 gum balls; 20 are yellow, 30 are blue, and 50 are red. The gum balls are well mixed inside the machine.
Jenny gets 10 gum balls from this machine.
What is your best prediction of the number that will be red?
Answer: 5
Explain why you chose this number.
**Item Scores**: 0 Minimum to 5 Maximum

<table>
<thead>
<tr>
<th>English Response</th>
<th>Native Language Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scorer 1: 4</td>
<td>Scorer 1: 4</td>
</tr>
<tr>
<td>Scorer 2: 4</td>
<td>Scorer 2: 3</td>
</tr>
<tr>
<td>Mean Score: 4</td>
<td>Mean Score: 3.5</td>
</tr>
</tbody>
</table>

**Commentary:**

The English sentence structure and spelling show some errors. The response shows the student understands the prompt and has a basic understanding of the concept in both languages. The Chinese grammar is in development.

Although the student arrives at the correct solution in both languages, the response in English provides better reasonings and explanations than the response in Chinese. The Chinese response contains no mathematical algorithms, just simple sentences acknowledging the larger number of red gumballs as the determining factor for the selection made. However, the student provides the same, correct number response in both languages.
Think carefully about the following question. Write a complete answer. You may use drawings, words, and numbers to explain your answer. Be sure to show all of your work.

2. The gum ball machine has 100 gum balls; 20 are yellow, 30 are blue, and 50 are red. The gum balls are well mixed inside the machine. Jenny gets 10 gum balls from this machine. What is your best prediction of the number that will be red?

Answer: 5

Explain why you chose this number.

I know all of them are 100, but red are 50, 30 are blue, 20 are yellow, the red were more, so it's like 20/100, 30/100, and 50/100. Fifty percent you will get a red gum, so I think that will be the best answer.
In both languages, the student provides a resourceful justification of the strategy used to solve the problem. The English response shows misuse of the rules of capitalization, repetition of prompt information, and the influence of Chinese sentence structures. Although the results are accurate in both languages, the student's reasoning and use of fractions as well as the organized sequential approach to solving the problem and reaching a conclusion are far more eloquent and complete in Chinese than English. However, the responses in both languages meet the criteria for the highest score.
Mathematics
Haitian-Creole
Item: Gumball Machine
Think carefully about the following question. Write a complete answer. You may use drawings, words, and numbers to explain your answer. Be sure to show all of your work.

The gum ball machine has 100 gum balls; 20 are yellow, 30 are blue, and 50 are red. The gum balls are well mixed inside the machine.

Jenny gets 10 gum balls from this machine. What is your best prediction of the number that will be red?

Answer:

5

Explain why you chose this number.

Because if there are 50 red, 30 blue and 20 yellow, she cannot get 10 red ones so I chose 5 red.
Substitution of words in the English version such as “are” for “or,” as well as overuse of words “so” and “and” are evidence of limited command of that language. Also, there are several letters omitted at the end of words, as well as plurals and capitalization.

Responses in Haitian-Creole do not restate the question posed. Therefore, the student uses the native word “Paske” to initiate the answer, which is the equivalent use of “because,” whenever a “why” question is addressed.

Although the English version lacks conventional grammar and spelling, the student uses the notion that 5 is half of 10, just as 50 is half of 100.

The solution to the problem is accurate in both language versions, although the explanation in Haitian-Creole is a bit more concise and clear. The response is scored consistently in both languages.
Mathematics
Spanish
Item: Caterpillar
A fourth-grade class needs 5 leaves each day to feed its 2 caterpillars. How many leaves would they need each day for 12 caterpillars?

**Response:**
I counted in fives.

**Translation from Native Language:**
I counted in fives.

**Calculations:**
12 caterpillars
60 leaves
5 leaves
2 caterpillars
I counted by fives.

**Analysis:**

- **Fluency in 1st Language:** 3
- **Language Proficiency:** 2
- **Teacher Rating:** 1

**Teacher Ratings**:

<table>
<thead>
<tr>
<th>Content Area: Math</th>
<th>Item: Caterpillar</th>
<th>Teacher Ratings*</th>
</tr>
</thead>
<tbody>
<tr>
<td>English, Spanish</td>
<td>Fluency in 1st Language</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>2nd Language Proficiency</td>
<td>2</td>
</tr>
<tr>
<td>Science</td>
<td>Math Fluency in 1st Language</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
</tr>
</tbody>
</table>

**Translation from Native Language:**
I counted in fives.

**Calculations:**
12 x 5 = 60
60 / 2 = 30
5 leaves
2 caterpillars
I counted by fives.

**Clipped Text:**
3. A fourth-grade class needs 5 leaves each day to feed its 2 caterpillars. How many leaves would they need each day for 12 caterpillars?

**Response:**
I counted by fives each day.

**Calculation:**
12 x 5 = 60
60 / 2 = 30
5 leaves
2 caterpillars
I counted by fives.

**Translation from Native Language:**
I counted in fives.

**Calculations:**
12 x 5 = 60
60 / 2 = 30
5 leaves
2 caterpillars
I counted by fives.

**Rating Scale:**

<table>
<thead>
<tr>
<th>Rating Scale</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Item Scores**: 0 Minimum to 3 Maximum

**English Response** | **Native Language Response**
---|---
Scorer 1: 1 | Scorer 1: 1
Scorer 2: 1 | Scorer 2: 1
Mean Score: 1 | Mean Score: 1

**Commentary:**

The student understands the word problem in both languages. The response is incorrect and consistent across languages. However, the problem-solving strategy is partially correct. There is evidence of grammar development in transition, with primary language influence by writing “I count in fives” vs. “I counted by fives.”

In the Spanish version, the student takes notes, creates a table, and makes drawings that help her construct her reasoning. The use of multiple algorithms and operations in her annotations shows that she reached the correct answer in the process, but transferred the incorrect response to the blank space provided. The raters scored the item consistently low in both languages since the number response is incorrect.
A fourth-grade class needs 5 leaves each day to feed its 2 caterpillars. How many leaves would they need each day for 12 caterpillars?

Answer: 30 leaves.

Use drawings, words, or numbers to show how you got your answer.

$$5 \div 2 = 2.5 \text{ leaves per one}$$
$$12 \times 2.5 = 30 \text{ leaves}$$

They needed 30 each day. Each day a caterpillar needs 5÷2 leaves. So we just times the answer by 12 and we can get the answer.
**Item Scores***:

<table>
<thead>
<tr>
<th>English Response</th>
<th>Native Language Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scorer 1: 3</td>
<td>Scorer 1: 3</td>
</tr>
<tr>
<td>Scorer 2: 3</td>
<td>Scorer 2: 3</td>
</tr>
<tr>
<td>Mean Score: 3</td>
<td>Mean Score: 3</td>
</tr>
</tbody>
</table>

* Item Scores: 0 Minimum to 3 Maximum

**Commentary:**

Although the solution and problem-solving strategies are correct for both problems, the response in Chinese is more elaborate. The student explains the reasoning behind the calculations in the Chinese version, whereas in the English response, she only uses algorithmic representations without justification or further elaboration.
A fourth-grade class needs 5 leaves each day to feed its 2 caterpillars. How many leaves would they need each day for 12 caterpillars?

**English Response:**

30 leaves.

12 ÷ 2 = 6
5 x 6 = 30

If we wanted to feed 12 caterpillars, we would need 30 leaves. Because 12 ÷ 2 = 6 times, 5 x 6 = 30 leaves.
<table>
<thead>
<tr>
<th>Item Scores: *</th>
<th>Notes:</th>
</tr>
</thead>
<tbody>
<tr>
<td>English Response</td>
<td>Native Language Response</td>
</tr>
<tr>
<td>Scorer 1: 3</td>
<td>Scorer 1: 3</td>
</tr>
<tr>
<td>Scorer 2: 3</td>
<td>Scorer 2: 3</td>
</tr>
<tr>
<td>Mean Score: 3</td>
<td>Mean Score: 3</td>
</tr>
</tbody>
</table>

* Item Scores: 0 Minimum to 3 Maximum

Commentary:
Although the solution is correct for both problems, the response in English is more elaborate because the student used graphic illustrations. The student uses the images to support his explanation. The response shows evidence of sentence structure influenced by the primary language. The Chinese response is also accurate, though it shows simplified writing.
A fourth-grade class needs 5 leaves each day to feed its 2 caterpillars. How many leaves would they need each day for 12 caterpillars?

**English Response:**

There are 12 caterpillars, the class needed 5 to feed two caterpillars. They needed 10 leaves to feed four caterpillars, they needed 15 leaves to feed six caterpillars, they needed 20 leaves to feed eight caterpillars, they needed 25 leaves to feed 10 caterpillars, and they needed 30 leaves to feed 12 caterpillars.

**Translation from Native Language:**

The answer is 30 leaves. Because we needed 10 leaves to feed 4 caterpillars and we needed 15 leaves to feed 6 caterpillars, we needed 25 leaves to feed 10 caterpillars, then we needed 30 leaves to feed 12 caterpillars.
**Commentary:**

The student is able to convey meaning and process in both languages.

The repeated use of the same sentence structure in English facilitates a sequential and logical structure for the student to solve the problem and explain its solution.

The responses in both languages use the same problem-solving approach and arrive at the correct answer. The student shows understanding of the concepts and skills needed to solve the problem.
A fourth-grade class needs 5 leaves each day to feed its 2 caterpillars. How many leaves would they need each day for 12 caterpillars?

Answer: 30 leaves.

Use drawings, words, or numbers to show how you got your answer.

2 \times 6 = 12 \text{ so it is 12 caterpillars, so I did } 5 \text{ leaves } \times 6 = 30 \text{ leaves.}
Item Scores:

<table>
<thead>
<tr>
<th>English Response</th>
<th>Native Language Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scorer 1: 3</td>
<td>Scorer 1: 3</td>
</tr>
<tr>
<td>Scorer 2: 3</td>
<td>Scorer 2: 3</td>
</tr>
<tr>
<td>Mean Score: 3</td>
<td>Mean Score: 3</td>
</tr>
</tbody>
</table>

* Item Scores: 0 Minimum to 3 Maximum

Commentary:

There is evidence of syntax development in the English version, as the subject/verb correspondence in the sentence is incorrect. Although the response in English is correct, the explanation is not as clear as in Chinese.

In both language versions the student arrives at a correct solution without a very elaborate explanation or justification for the strategy chosen.
### Teacher Ratings*

<table>
<thead>
<tr>
<th>Fluency in 1st Language</th>
<th>Literacy in 1st Language</th>
<th>Fluency in 2nd Language</th>
<th>Literacy in 2nd Language</th>
<th>Science Proficiency</th>
<th>Math Proficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

---

A fourth-grade class needs 5 leaves each day to feed its 2 caterpillars. How many leaves would they need each day for 12 caterpillars?

**Answer:**

Use drawings, words, or numbers to show how you got your answer.

Because the caterpillars each day need to eat 2 and half leaves, 12 caterpillars need 2 and half times 12, the answer is 30.

So 30 leaves to 12 caterpillars.
Item Scores*:

<table>
<thead>
<tr>
<th>English Response</th>
<th>Native Language Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scorer 1: 3</td>
<td>Scorer 1: 3</td>
</tr>
<tr>
<td>Scorer 2: 3</td>
<td>Scorer 2: 3</td>
</tr>
<tr>
<td>Mean Score: 3</td>
<td>Mean Score: 3</td>
</tr>
</tbody>
</table>

* Item Scores: 0 Minimum to 3 Maximum

Commentary:

The Chinese response is complete, accurate, and clear. The student demonstrates understanding of the operational solution to the problem in a sequential and thorough manner. The English version uses the same strategy, but the use of language is limited and emergent English sentence structure and vocabulary development are evident. This can be seen in the student's use of "mine" vs. "times."

Despite these differences, the student provides correct responses in both languages, which are consistently scored high.
Mathematics
Haitian-Creole
Item: Caterpillar
A fourth-grade class needs 5 leaves each day to feed its 2 caterpillars. How many leaves would they need each day for 12 caterpillars?

Answer:

Student used drawing to answer the question.

Translation from Native Language:
I say $12 \times 5 = 60$.

Response:

Use drawings, words, or numbers to show how you got your answer.
### Item Scores:

<table>
<thead>
<tr>
<th>English Response</th>
<th>Native Language Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scorer 1: 2</td>
<td>Scorer 1: 1</td>
</tr>
<tr>
<td>Scorer 2: 1</td>
<td>Scorer 2: 1</td>
</tr>
<tr>
<td>Mean Score: 1.5</td>
<td>Mean Score: 1</td>
</tr>
</tbody>
</table>

* Item Scores: 0 Minimum to 3 Maximum

### Commentary:

The responses in both languages are incorrect. However, the ways in which the student provides her responses vary with each language.

The student uses illustrations and the same representation strategies in both language versions, but in the English version she does not use words to show her work. In the Haitian–Creole version, the student formalizes her reasoning with an operational algorithm.
Appendix A
Suggested Fluency and Literacy Rating Scales

The scale in this Appendix is different from that used by the teachers. However, we included a scale in the appendix for the purpose of helping participants in a professional development activity to reflect on how the language quality in the responses reflect what the rubrics describe.

Fluency in 1st language

5 = Maximum. Demonstrates full understanding expected of a native speaker of the same age. Able to accurately use a broad range of syntactic features, and has a broad vocabulary.

4 = Gaining. Understands all but some advanced structures. Uses complex structures with only occasional errors in syntax. Vocabulary is varied.

3 = Developing. Understands simplified speech with some repetition and rephrasing. Uses complex structures, but over-generalizes rules of grammar.

2 = Beginning. Understands slow simple speech and requires repetition. Uses simple speech patterns with limited vocabulary.

1 = Minimum. Understands little or none of the native language.

Literacy in 1st language

5 = Maximum. Can read and write well enough to meet content-area expectations in the native language.

4 = Gaining. Can read most grade level materials, but needs considerable help with academic vocabulary. Can write sentences with some grammatical accuracy.

3 = Developing. Understands printed material and can retell familiar stories. Writes using high-frequency words and repetitive sentence structures. Can supply simple information and spelling is partially correct.

2 = Beginning. Knows English alphabet/sounds and can read simple patterns and read from language experience stories. Spelling is often nonstandard and may be influenced by native language letter-sound system. Can produce copied pattern sentences.

1 = Minimum. No functional ability to read or write in English.

Fluency in 2nd language

5 = Maximum. Demonstrates full understanding expected of a English-fluent speaker of the same age. Able to accurately use a broad range of syntactic features and has a broad vocabulary.

4 = Gaining. Understands all but some advanced structures. Uses complex structures with only an occasional error in syntax. Vocabulary is varied.

3 = Developing. Understands simplified speech with some repetition and rephrasing. Uses complex structures, but over-generalizes English rules of grammar.

2 = Beginning. Understands slow simple speech and requires repetitions. Uses simple speech patterns with limited vocabulary.

1 = Minimum. Understands little or no English except for a word or two.

Literacy in 2nd language

5 = Maximum. Can read and write well enough to meet content-area expectations in English.

4 = Gaining. Beginning to read grade-level material, but needs considerable help with academic vocabulary. Can write sentences with some grammatical accuracy.

3 = Developing. Understands printed material and can retell familiar stories. Writes using high-frequency words and repetitive sentence structures. Can supply simple information and spelling is partially correct.

2 = Beginning. Knows English alphabet/sounds and can read simple patterns and read from language experience stories. Spelling is often nonstandard and may be influenced by native language letter-sound system. Can produce copied pattern sentences.

1 = Minimum. No functional ability to read or write in English.
Appendix B

NAEP Score of Student's Written Performance in Science: Respiratory System of Tadpoles

Student Identification No. ____________________________

Name of Assessment ____________________________

Score(s) Identification No. ____________________________

Language of Assessment ____________________________

Directions: Circle the score that best matches the student's performance.

3 = Complete: Student describes both the source of oxygen and the organs used for breathing in tadpoles.

2a = Partial: Student correctly describes the source of oxygen (e.g., "tadpoles get oxygen from water")

2b = Partial: Student correctly describes the organs used for respiration for tadpoles.

1 = Unsatisfactory/Incorrect: Student is unable to correctly identify the source of oxygen (water), or the organs used for respiration (gills) for tadpoles.

Rationale for Score
(Use the other side of this sheet if more space is needed to record your comments)

What is your rationale for the NAEP science score?

1.

2.

3.
### NAEP Score of Student's Written Performance in Science: Heat

**Student Identification No.** ____________________________  
**Score(s) Identification No.** ____________________________  
**Name of Assessment** ____________________________  
**Language of Assessment** ____________________________

**Directions:** Circle the score that best matches the student's performance.

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td><strong>Complete:</strong> Student demonstrates an understanding that heat is a form of energy that can be produced when one form of energy is transformed to another form of energy.</td>
</tr>
<tr>
<td>2</td>
<td><strong>Partial:</strong> Student response shows some understanding by saying that the warmth is caused by friction, but an incorrect or no explanation is given.</td>
</tr>
<tr>
<td>1</td>
<td><strong>Unsatisfactory/Incorrect:</strong> Student response does not indicate any understanding of why a nail becomes warm when it is hammered into a piece of wood.</td>
</tr>
<tr>
<td>1</td>
<td><strong>Unsatisfactory/Incorrect:</strong> Student response does not indicate any understanding saying that the warmth is caused by friction, but an incorrect or no explanation is given.</td>
</tr>
</tbody>
</table>

**Rationale for Score**  
(*Use the other side of this sheet if more space is needed to record your comments.*)

What is your rationale for the NAEP science score?

1.  
2.  
3.  

**CCSSO LEP Work Sampler Booklet**  
182
# NAEP Score of Student's Written Performance in Science: Erosion

Student Identification No. _________________________  
Score(s) Identification No. _________________________  
Name of Assessment _______________________________  
Language of Assessment ____________________________  

**Directions:** Circle the score that best matches the student's performance.

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td><strong>Complete:</strong> Student chooses picture A and explains why referring to both the mountain and the river.</td>
</tr>
<tr>
<td>2a</td>
<td><strong>Partial:</strong> Student chooses picture A and explains why referring to the mountains only.</td>
</tr>
<tr>
<td>2b</td>
<td><strong>Partial:</strong> Student chooses picture A and explains why referring to the river only.</td>
</tr>
<tr>
<td>1</td>
<td><strong>Unsatisfactory/Incorrect:</strong> Student chooses neither picture or chooses A or B, but does not explain why.</td>
</tr>
</tbody>
</table>

**Credited responses include:** Picture A, Mountains (rounded, smaller, worn down, smother) River (many curves or bends, bigger or wider)

## Rationale for Score

(Use the other side of this sheet if more space is needed to record your comments)

What is your rationale for the NAEP science score?

1.  
2.  
3.  

---
## NAEP Score of Student's Written Performance in Mathematics: Figures

**Student Identification No.** ___________________________  
**Score(s) Identification No.** ___________________________

**Name of Assessment** ___________________________  
**Language of Assessment** ___________________________

**Directions:** Circle the score that best matches the student's performance.

### Scoring guide:

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Student gives at least two reasons why they are alike and at least two reasons why they are different (two alike reasons re not both &quot;a&quot;)</td>
</tr>
<tr>
<td>4</td>
<td>Student gives two reasons why figures are alike and one reason why they are different or one reason why they are alike and two reasons why they are different</td>
</tr>
<tr>
<td>3</td>
<td>Student gives one correct reason alike and one correct reason different or two reasons alike or two reasons different</td>
</tr>
<tr>
<td>2</td>
<td>A nonspecific response (i.e., the one on the right is skinner) or only one correct reason (alike or different)</td>
</tr>
<tr>
<td>1</td>
<td>Incorrect response</td>
</tr>
</tbody>
</table>

### Solution #1:

*The figures are alike because:*

1. They both have 4 sides (or 4 corners or 4 angles)
2. They both have parallel sides
3. They both have two sets of sides that are the same length
4. They have the same area
5. They have the same length (Base)
6. They have the same height
7. They both have little squares

*Note:*  
Do not accept: They both have lines that are straight. Also 4 sides and 4 angles not considered different reasons.

### Solution #2:

*The figures are different because:*

8. One has 4 equal angels and the other does not
9. One has right angles or perpendicular lines and the other does not (students don't need to make the comparison, i.e., they can just say "one has 4 equal angels")
10. One is "Slanter" than the other (or one takes up full squares and the other does not)
11. They have different perimeters.

*Note:*  
Do not accept "They're not both the same shape."

### Rationale for Score

*(Use the other side of this sheet if more space is needed to record your comments)*

What is your rationale for the NAEP mathematics score?

1. 
2. 
3. 
4. 
5.
NAEP Score of Student's Written Performance in Mathematics: Gum Ball Machine

Student Identification No. ____________________________  Score(s) Identification No. ____________________________
Name of Assessment ________________________________  Language of Assessment ________________________________

Directions: Circle the score that best matches the student's performance.

<table>
<thead>
<tr>
<th>Scoring guide:</th>
<th>Solution:</th>
<th>*Numerical Answer</th>
<th>Possible Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 = Correct response (4-6 gum balls, half of the gum balls are red)</td>
<td>5 gum balls would probably be red. Half of the 100 gum balls in the machine are red, so half of what she gets out should be red. Answers such as 4 or 6 are acceptable if explanation is correct. A less correct explanation: There are more reds in the machine so more reds would come out. Also acceptable: 5 because 10 percent of 50 is 5 and 10 is 10 percent of 100) OR 5 because 50 divided by 10 is 5</td>
<td>1</td>
<td>1 or 3</td>
</tr>
<tr>
<td>4 = Answers 4-6 gum balls with explanation that there are more red gum balls than any other color</td>
<td></td>
<td>2</td>
<td>1 or 3</td>
</tr>
<tr>
<td>3 = Answers 3-7 with no explanation or insufficient explanation OR Answers 1 or 2 or 8 or 10 gum balls with explanation that there are more red gum balls</td>
<td></td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>2 = Answers greater than 10 or no number with explanation that there are more red gum balls</td>
<td></td>
<td>4,5,6</td>
<td>3, 4, or 5</td>
</tr>
<tr>
<td>1 = Incorrect response</td>
<td></td>
<td>7</td>
<td>3</td>
</tr>
</tbody>
</table>

*Non-integer answer treated as a range (e.g. 3.5 is scored as 3 or 4). Range scored according to worst answer (e.g. 3 to 4 scored as 3).
NAEP Score of Student's Written Performance in Mathematics: Caterpillars

Student Identification No. __________________________ Score(s) Identification No. __________________________
Name of Assessment ______________________________ Language of Assessment __________________________

Directions: Circle the score that best matches the student's performance.

<table>
<thead>
<tr>
<th>Scoring guide:</th>
<th>Solution:</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 = Correct answer with work correct and complete method or process shown.</td>
<td>The answer is 30</td>
</tr>
<tr>
<td>2 = Correct answer with no work shown (including work) or correct method with computational error (including correct method with answer 60)</td>
<td>12 caterpillars is 6 pairs; 5 leaves for each pair</td>
</tr>
<tr>
<td>1 = Incorrect response, including 60 (with or without work) if work → 60</td>
<td>$5 \times 6 = 30$ or $12 \div 30$ or pictorial response are also possible</td>
</tr>
</tbody>
</table>

Rationale for Score
(Use the other side of this sheet if more space is needed to record your comments)

What is your rationale for the NAEP mathematics score?

1. 

2. 

3. 

---

127 caterpillars is 6 pairs; 5 leaves for each pair. The solution is 30.
Two main analyses were conducted. First, we tested the statistical significance of the item score differences across languages within each native language group. Second, we performed a series of generalizability studies to examine how much of the score variation within each cultural group was due to the main and interaction effect of four sources of score variability: student, item, language, and rater.

### Mean Score Differences

Table 6 shows the mean scores obtained by the students from each cultural group on each item when they were tested in their native languages and in English. A series of t-tests revealed that only about one third of the mean score differences were statistically significant (Table 7). Some of these differences are in favor of the students' native language for some items and in favor of English for other items. The set of items on which students perform better in their native language than in English is different for each group. That is the case even for the two classes of Chinese native speakers that participated in the study.

<table>
<thead>
<tr>
<th>Item</th>
<th>N</th>
<th>Spanish</th>
<th>English</th>
<th>N</th>
<th>Haitian Creole</th>
<th>English</th>
<th>N</th>
<th>Chinese</th>
<th>English</th>
<th>N</th>
<th>Chinese</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tadpoles</td>
<td>18</td>
<td>1.9167</td>
<td>1.9167</td>
<td>21</td>
<td>1.1905</td>
<td>1.0952</td>
<td>26</td>
<td>1.6731</td>
<td>1.7500</td>
<td>28</td>
<td>0.8929</td>
<td>0.6429</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.6243)</td>
<td>(.7326)</td>
<td></td>
<td>(.3345)</td>
<td>(.3008)</td>
<td></td>
<td>(.4678)</td>
<td>(.5701)</td>
<td></td>
<td>(.7741)</td>
<td>(.6647)</td>
</tr>
<tr>
<td>Frogs</td>
<td>18</td>
<td>1.6111</td>
<td>1.4167</td>
<td>21</td>
<td>1.0952</td>
<td>1.2143</td>
<td>26</td>
<td>1.6731</td>
<td>1.4615</td>
<td>28</td>
<td>0.7679</td>
<td>0.5893</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.5016)</td>
<td>(.8787)</td>
<td></td>
<td>(.4068)</td>
<td>(.5825)</td>
<td></td>
<td>(.4887)</td>
<td>(.5277)</td>
<td></td>
<td>(.6733)</td>
<td>(.6391)</td>
</tr>
<tr>
<td>Erosion</td>
<td>18</td>
<td>1.3333</td>
<td>1.3333</td>
<td>21</td>
<td>0.9286</td>
<td>1.1667</td>
<td>26</td>
<td>1.6538</td>
<td>1.5769</td>
<td>28</td>
<td>1.1786</td>
<td>1.0536</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.5688)</td>
<td>(.5688)</td>
<td></td>
<td>(.5312)</td>
<td>(.3979)</td>
<td></td>
<td>(.8691)</td>
<td>(.7575)</td>
<td></td>
<td>(.8630)</td>
<td>(.7857)</td>
</tr>
<tr>
<td>Heat</td>
<td>18</td>
<td>1.0883</td>
<td>1.0833</td>
<td>21</td>
<td>0.9762</td>
<td>1.0476</td>
<td>26</td>
<td>0.9231</td>
<td>1.1346</td>
<td>28</td>
<td>1.3036</td>
<td>0.6071</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.3536)</td>
<td>(.3930)</td>
<td></td>
<td>(.8871)</td>
<td>(.7891)</td>
<td></td>
<td>(.1840)</td>
<td>(.4137)</td>
<td></td>
<td>(.8426)</td>
<td>(.5830)</td>
</tr>
<tr>
<td>Figures</td>
<td>20</td>
<td>2.5250</td>
<td>2.3000</td>
<td>23</td>
<td>2.2609</td>
<td>3.0217</td>
<td>26</td>
<td>2.3846</td>
<td>2.5577</td>
<td>31</td>
<td>2.5323</td>
<td>1.1290</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.1177)</td>
<td>(1.1517)</td>
<td></td>
<td>(1.1167)</td>
<td>(1.4653)</td>
<td></td>
<td>(1.4234)</td>
<td>(1.6083)</td>
<td></td>
<td>(1.7413)</td>
<td>(1.2515)</td>
</tr>
<tr>
<td>Gumball</td>
<td>20</td>
<td>1.7250</td>
<td>2.0000</td>
<td>23</td>
<td>2.6304</td>
<td>2.8043</td>
<td>26</td>
<td>3.0769</td>
<td>3.2885</td>
<td>31</td>
<td>2.9516</td>
<td>1.6290</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.5730)</td>
<td>(.9319)</td>
<td></td>
<td>(1.4711)</td>
<td>(1.2036)</td>
<td></td>
<td>(1.1462)</td>
<td>(1.1591)</td>
<td></td>
<td>(1.8590)</td>
<td>(1.6531)</td>
</tr>
<tr>
<td>Caterpillar</td>
<td>20</td>
<td>1.5750</td>
<td>1.6250</td>
<td>23</td>
<td>0.7391</td>
<td>1.2609</td>
<td>26</td>
<td>2.0962</td>
<td>2.1154</td>
<td>31</td>
<td>2.1774</td>
<td>1.5968</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.8472)</td>
<td>(.9014)</td>
<td></td>
<td>(.4490)</td>
<td>(.8774)</td>
<td></td>
<td>(.9800)</td>
<td>(.9829)</td>
<td></td>
<td>(1.1442)</td>
<td>(1.3688)</td>
</tr>
</tbody>
</table>
Table 7. Score differences for items administered in the students' native languages and in English. Positive and negative values indicate, respectively, a difference in favor of the native language and English. Statistical significance (p < .05) indicated with an asterisk.

<table>
<thead>
<tr>
<th>Tadpole</th>
<th>Frog</th>
<th>Earth</th>
<th>Heat</th>
<th>Figures</th>
<th>Gumball</th>
<th>Caterpillar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spanish</td>
<td>0</td>
<td>.194</td>
<td>0</td>
<td>0</td>
<td>.225</td>
<td>-.275</td>
</tr>
<tr>
<td>Haitian</td>
<td>.095*</td>
<td>-.119</td>
<td>-.238*</td>
<td>-.071</td>
<td>-.760*</td>
<td>-.173</td>
</tr>
<tr>
<td>Chinese 1</td>
<td>-.076</td>
<td>.211</td>
<td>.076</td>
<td>-.211*</td>
<td>-.173</td>
<td>-.211</td>
</tr>
<tr>
<td>Chinese 2</td>
<td>.250</td>
<td>.178</td>
<td>.125</td>
<td>.696*</td>
<td>1.403*</td>
<td>1.322*</td>
</tr>
</tbody>
</table>

Generalizability Studies

We used generalizability (G) theory (Cronbach, Gleser, Nanda, & Rajaratnam, 1972; Shavelson & Webb, 1991) to assess how much scores varied due to different sources of score variation. For each native language group (Spanish, Chinese, and Haitian-Creole) and within each content area (science and mathematics) we performed a series of G studies with a student (p) x item (i) x language (l) x rater (r) design, which enabled us to estimate the relative magnitude of the effect of these sources of score variation and their interactions. In these analyses, student, item, and rater were treated as random, whereas language was treated as fixed.

Consistent patterns of score variation were observed for the three native language groups in separate analyses. The results are summarized in Table 8.

Table 8. Percentages of score variation for science and mathematics (averaged across native language groups).

<table>
<thead>
<tr>
<th>Source of Score Variation</th>
<th>Science</th>
<th>Mathematics</th>
</tr>
</thead>
<tbody>
<tr>
<td>student (p)</td>
<td>8</td>
<td>14</td>
</tr>
<tr>
<td>item (i)</td>
<td>14</td>
<td>4</td>
</tr>
<tr>
<td>language (l)</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>rater (r)</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>si</td>
<td>20</td>
<td>21</td>
</tr>
<tr>
<td>sl</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>sr</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>il</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>ir</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>ir</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>sil</td>
<td>23</td>
<td>23</td>
</tr>
<tr>
<td>sir</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>slr</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>slr,e</td>
<td>16</td>
<td>19</td>
</tr>
</tbody>
</table>

Note: Percentages do not necessarily add up to 100 due to rounding.

Four facts stand out. First, the relative magnitude of the effect of each source of score variation is similar across science and mathematics. The only exception is the main effect of item, which produces 14 percent of the score variation for science but only 4 percent of the score variation for mathematics. Student scores across items vary more for science than for mathematics.

Second, score variation due to rater and the interaction of rater and other facets (sr, ir, lir, sir, slr, ilr) is small. Raters are not an important source of measurement error.

Third, consistent with evidence reported in the literature (e.g., Ruiz-Primo, Baxter, & Shavelson, 1993; Shavelson, Baxter, &
the interaction of student and item (si) is a considerable source of score variation. A student who performs well on one item does not necessarily perform well on another item from the same content area.

Fourth, the largest score variation was produced by the interaction of student, item, and language (sil) (23 percent for both science and mathematics) whereas the relative magnitude of score variation due to the main effect of language (2 percent for science and 6 percent for mathematics) is small. Given the goals of this project, these are the most important results. The language chosen for a test does not produce itself any considerable effect on the performance of English learners. English learners perform well depending on both item and language. A given student may perform well for some items administered in his or her native language but not as well for other items administered also in his or her native language. This same student's performance on those other items may be better if they are administered in English.
Appendix D

References


Council of Chief State School Officers (Fall, 1999). Data from the Annual Survey: Student Assessment Program. CCSSO Washington, DC.


Appendix E

SAMPLER DEVELOPMENT COMMITTEE

Amien Lau
Daniel Webster
Elementary School
San Francisco, CA

Elizabeth Pacheco
Dr. LA Nixon Elementary School
El Paso, Texas

Marcelle Poulos
Marshall Elementary School
San Francisco, CA

Christine Damas
Community School District 17
Brooklyn, New York

Marie Joseph
P.S. 22
Brooklyn, New York

Rosie Chow
Marshall Elementary School
San Francisco, CA

Shouben Li
McCall Elementary School
Philadelphia, PA

Rick Jean Marie
PS 189
Brooklyn, New York

CONSULTANTS
Guillermo Solano-Flores
WestEd
California

Ursula Sexton
WestEd
California

Cecilia Navarrete
Independent Consultant
Albuquerque, New Mexico

REVIEWERS
Aldo Aguirre
Nevada Department of Education

Laura Ayala
Texas Education Agency

Oona Cheung
CCSSO

Chopin Kiang
Nevada Department of Education

John Olson
CCSSO

Christopher Renner
Kansas State Department of Education

Maria Medina Seidner
Texas Education Agency

Jeanette Spencer
California Department of Education

John Tai
Formerly with CCSSO

346
NOTICE

Reproduction Basis

☐ This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☐ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").

EFF-089 (3/2000)