ASSESSMENT AND EXERCISE SPECIFICATIONS

2003 NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS (NAEP) IN

FOREIGN LANGUAGE

May 2000

Developed Under Contract Number ED99CO0102 for the
National Assessment Governing Board by the
Center for Applied Linguistics, American Council on the Teaching of Foreign Language,
and the American Institutes for Research

TABLE OF CONTENTS

INTRODUCTION 3

CHAPTER 1. ASSESSMENT SPECIFICATIONS FOR THE SPANISH LANGUAGE ASSESSMENT 5

Student Sampling 5

Construction of the Assessment 9

Construction of the Spanish Language Assessment Exercise Pool 16

Review Process and Criteria 20

Administration Specifications 22

Scoring and Reporting Specifications 23

Recommendations for Special Studies 27

CHAPTER 2. EXERCISE SPECIFICATIONS FOR THE SPANISH LANGUAGE ASSESSMENT 29

General Characteristics of Exercise Development 29

Preliminary Achievement Level Descriptions 30

Preliminary Recommendations for Achievement Levels 33

Exercise Format and Content Specifications 37

Specifications for Scoring Guides 53

Rater Training Specification 55

Chapter 3. LANGUAGE SURVEY/ BACKGROUND QUESTIONNAIRE 57

Background to the LS/BQ 57

Administration of the LS/BQ 57

Student Variables 58

Teacher and Instructional Variables 59

School and Community Variables 59

REFERENCES 61

APPENDIX A. SAMPLE EXERCISES FOR SPANISH LANGUAGE ASSESSMENT 63

Interpretive Mode: Sample Listening-Based Exercise 64

Interpretive Mode: Sample Reading-Based Exercise 65

Presentational Mode: Sample Writing-Based Exercise 66

Interpersonal Mode: Sample Conversation-Based Exercises 67

APPENDIX B. Participants in NAEP Foreign Language Consensus Building Process **75**

The Consensus Steering Committee **75**

The Consensus Planning Committee **76**

The Technical Advisory Panel **77**

The Project Management Team **77**

LIST OF TABLES

Table 1. Categories of Students 6, 26
Table 2. Areas of Communication to be Assessed 16
Table 3. Specifications for Exercise and Response Formats 17
Table 4. Achievement Level Policy Definitions 30
Table 5. Interpretive Mode: Reading-Based Tasks 33
Table 6. Interpretive Mode: Listening-Based Tasks 34
Table 7. Presentational Mode: Writing-Based Tasks 35
Table 8. Interpersonal Mode: Conversation-Based Tasks 36

LIST OF FIGURES

Figure 1: Stages 1 and 2 of the Spanish NAEP Sampling Procedure 8
Figure 2. The Foreign Language NAEP Assessment Framework 9
Figure 3. The Spanish NAEP Assessment Framework 15
Figure 4. Distribution of Exercises Across Two Levels of the Assessment 19
Figure 5. Stages of Follow-Up Questions 53

INTRODUCTION

The National Assessment Governing Board (NAGB) has targeted the year 2003 for the first foreign language NAEP (FL NAEP). In May 1999, NAGB awarded a contract to the Center for Applied Linguistics (CAL) to conduct a national consensus building project. CAL worked in collaboration with the American Council on the Teaching of Foreign Languages (ACTFL) and the American Institutes for Research (AIR) to develop recommendations for the Governing Board on the framework and specifications for the FL NAEP.

The FL NAEP will provide information to the nation on how well students in the United States can communicate in languages other than English. Specifically, results from the 2003 assessment will report on how well representative samples of 12th grade students who have learned Spanish in a variety of ways and for different lengths of time can communicate in Spanish.

The Spanish NAEP will be based on the consensus building committees' general framework for assessing communicative ability in languages other than English. In this framework, listening, speaking, reading and writing skills are assessed within three modes of communication. The *interpersonal* mode involves two-way, interactive communication, such as conversing face-to-face or exchanging e-mail messages. The *interpretive* mode relates to the one-way understanding of spoken or written language, such as listening to a broadcast or reading a magazine. The *presentational* mode involves creating one-way spoken or written communication, such as giving a speech or writing a story.

The framework for the foreign language NAEP calls for communicative ability to be assessed through authentic communication tasks as called for in daily life, school, and work. The assessment tasks will reflect four interrelated goals that provide the basis for communication. The four goals include:

- gaining knowledge of other **cultures**;
- **connecting** with other academic subject areas to acquire knowledge;
- developing insights into the nature of language and culture through **comparisons**; and
- participating in multilingual **communities** at home and around the world.

Performances on assessment tasks will be evaluated on the criterion of how well the student understands (comprehension) and can be understood (comprehensibility). This criterion subsumes language knowledge, the appropriate use of communication strategies (such as asking for clarification or inferring the meaning of unknown words from context), and the application of cultural knowledge to enhance communication.

The Spanish NAEP for 2003 will be based on the general framework described above. However, the consensus building committees recommend that the Spanish NAEP focus on assessing four of the six assessment areas in the general FL NAEP framework. The Spanish assessment will require demonstration of

- listening and speaking in the interpersonal mode;
- listening in the interpretivc mode;
- reading in the interpretive mode; and
- writing in the presentational mode.

The two assessment areas not assessed by the Spanish NAEP, due to practical considerations of time and expense, are reading and writing in the interpersonal mode and speaking in the presentational mode. The four assessment areas chosen are those most used in real-world communication by secondary school students.

The 2003 Spanish NAEP will assess students' communicative ability in Spanish using a variety of oral and written stimulus materials, as well as other nontextual materials such as photos, artwork, graphics, and videos. Different response modes will be combined with the stimuli in developing items. The response modes will be single correct option multiple-choice items; written constructed response (open-ended) exercises, both short answer and extended response; and oral performance exercises (conversation). The following pages present detailed specifications for the overall assessment and for the exercises (items).

An Item Writers' Guide that presents basic rules for good item construction for all item formats is to be provided by the test development contractor. The Guide should include criteria for developing items using a combination of response modes and stimulus materials and must conform to the specifications set forth in this document, the *National Assessment Governing Board (NAGB) Policy on Cognitive Item Development and Review*, and any formatting requirements of NAEP.

This document is intended for the test developer and is divided into two main sections: 1) Assessment Specifications and 2) Exercise Specifications. The Assessment Specifications section provides an overall description of the sampling plan for the assessment and specifies the construction, review, administration, and scoring of the assessment. The Exercise Specifications section presents the preliminary Achievement Level descriptions and specifies the construction of the exercises in terms of format and content. Specifications also are provided in this section for designing scoring rubrics and for the selection and training of raters. Sample exercises for the Spanish Language Assessment are included in Appendix A. Members of the consensus building committees and the Technical Advisory Panel are presented in Appendix B.

CHAPTER 1. ASSESSMENT SPECIFICATIONS FOR THE SPANISH LANGUAGE ASSESSMENT

Student Sampling

The consensus building committees recommend a two-stage sampling procedure for the Spanish NAEP:

- Stage 1: Language Survey/Background Questionnaire administered to a nationally representative sample of 12th graders
- Stage 2: Spanish Language Assessment administered to a representative sample of the national subpopulation of Spanish language learners in grade 12

In the first stage, a web based Language Survey/Background Questionnaire (LS/BQ) will be administered to a nationally representative sample of 12th graders sufficiently ahead of the Spanish assessment to ensure that the data collected are analyzed in time to inform the second stage of the sampling. This sample will contain learners of Spanish; learners of other foreign languages; learners of no foreign language; heritage language learners (students who are learning a foreign language to which they have access through their home environment); and speakers of a foreign language as a native or heritage language who do not have U.S. school learning experiences in the language. All students will be asked questions about their experiences with foreign language learning as well as asked to give a self-report on their communicative abilities in languages other than English.

The second stage will be the administration of the Spanish NAEP to a representative sample of the national subpopulation of Spanish language learners in grade 12. Students participating in stage two will be drawn from the larger sample of stage one students.[1] The course-taking results of the LS/BQ survey, verified by school records where possible, will be used to generate the sample for the main Spanish Language Assessment (Spanish NAEP). The sample of twelfth grade students who take the main assessment will include:

- those who are currently studying or who have studied Spanish
- those who acquired Spanish language proficiency outside of an academic setting.

Students with School Learning Experience

As explained in the Framework document, the critical issue to be examined by the foreign language NAEP is "how long does it take students to reach meaningful levels of achievement in a foreign language?" To examine the relationship between length of study and foreign language achievement, it is necessary to include in the Spanish NAEP large enough representative samples of 12th grade students to report achievement outcomes by varying lengths of time. Student samples also will need to be classified by whether or not they are currently studying the language.

[1] The first stage sample must contain representative samples of subgroups targeted in the second stage.

In addition, there are students in the United States for whom Spanish is a heritage language; that is, a language to which they have access outside of the school classroom. Heritage language students who participate in the Spanish NAEP (Stage 2) may or may not have received formal academic instruction in Spanish. Those heritage language speakers who have learned Spanish in school may have studied it in courses tailored to non-heritage language speakers. Alternatively, they may be participating in the growing number of courses titled "Spanish for Spanish Speakers." The consensus building committees recognize that it is difficult to categorize heritage language learners adequately by numbers of years of study. Thus, they recommend reporting achievement of heritage language learners who have studied Spanish in a U.S. school as a separate group.

For this reason, the sample of students *with* school learning experience will be defined by years of study and will include both heritage and non-heritage language learners, although the results will be reported separately. Results for students without school learning experience also will be reported. Table 1 shows the groups of students for whom achievement should be reported. Representative samples of students will be required for each of these groups.

Table 1. Categories of Students

1. Non-Heritage Language Students With U.S. School Experience Learning Spanish, Currently Studying Spanish in 12th Grade	
1A	5 or more years of continuous study (i.e., study was begun in 8th grade or earlier)
1B	4 years of continuous study (i.e., study was begun in 9th grade)
1C	3 years of continuous study (i.e., study was begun in 10th grade)
1D	2 years of continuous study (i.e., study was begun in 11th grade)
1E	1 year of study (i.e., study was begun in 12th grade)
2. Non-Heritage Language Students With U.S. School Experience Learning Spanish, Who Have Completed Their Last Spanish Course Prior to Current Year	
2A	3 or more years of study
2B	2 years of study
2C	1 year of study
3. Spanish Heritage Language Speakers with U.S. School Spanish Learning Experience	
4. Spanish Speakers (Heritage and Non-Heritage) without U.S. School Spanish Learning Experience	

As stated above, representative samples of each of the ten groups identified in Table 1 will be assessed with the Spanish NAEP in stage 2 of the study. Results in terms of NAEP achievement levels will be reported for each group, since the focus of the report will be on the connection between length of foreign language study and achievement.

The sampling of schools chosen to participate in the FL NAEP will need to be done very carefully so that each group for which results are reported are valid representations of the entire population for that group. The need for representative samples means that students with 5 or more years of continuous study may need to be oversampled to ensure that valid and reliable observations can be made about them. If the numbers of students in 1A are large enough, data

may be reported for subgroups, such as students with more than 6 years of continuous study (i.e., continuous study beginning prior to 7th grade) in order to provide more information about the relationship between length of study and achievement.

An issue for sampling is that the schools will be selected before the students are selected, so any over sampling of students will have to be within the selected schools. Traditional NAEP sampling methods might not provide enough of some types of students if they are clustered in certain types of schools, and few of those schools are selected. The Technical Advisory Panel recommends that the sampling contractor use multiple sources of data, e.g., the NAEP Transcript Study, AP course data, the ACTFL 2000 survey, the 2002 field test, and the 2000 Census, to discover course-taking patterns and other information for identifying schools where the critical populations exist. If schools with a higher probability of populations of students needing to be over sampled exist and can be identified before school sampling, the sampling contractor should select these schools.

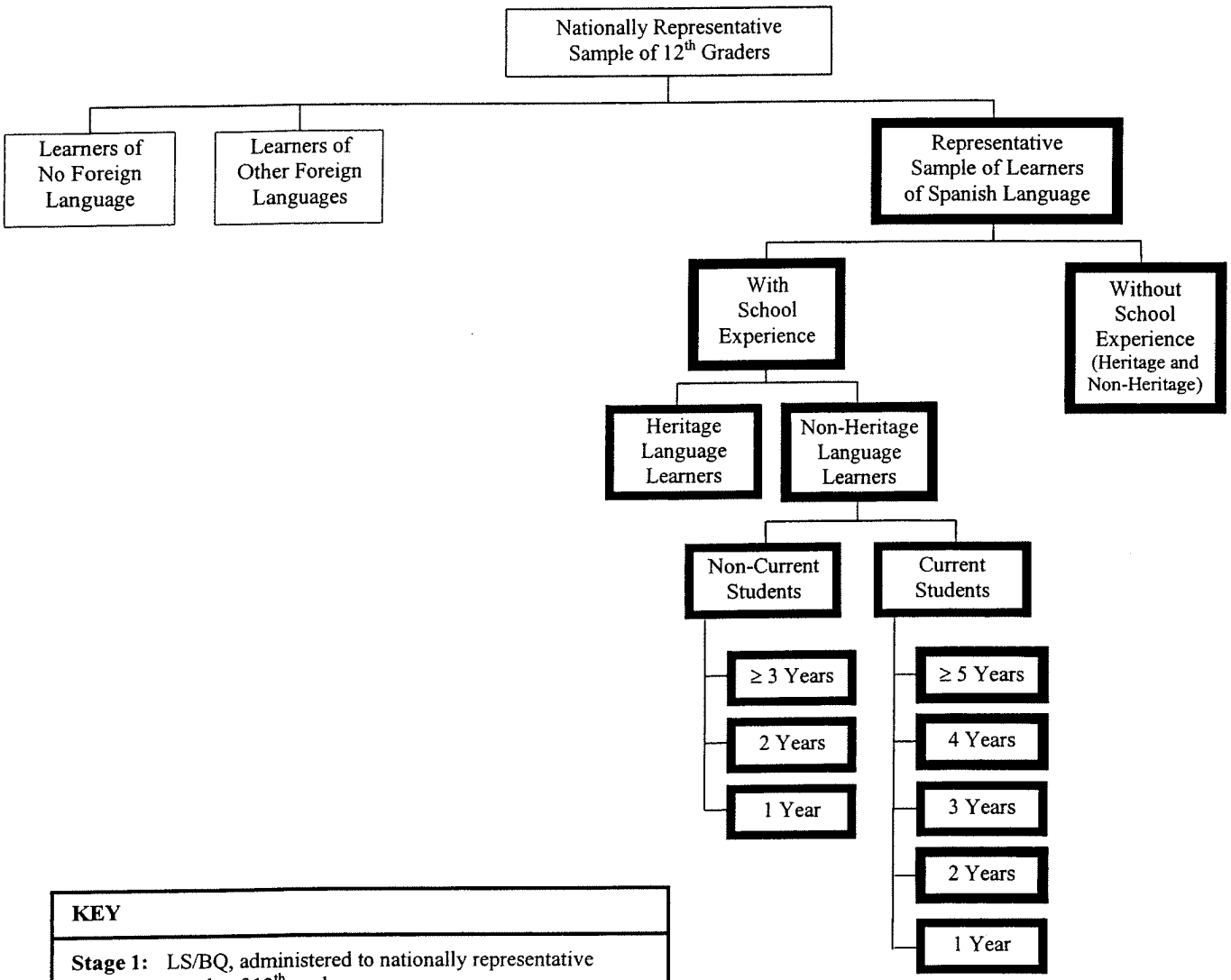
The sampling contractor will need to be informed by sources like those listed above in order to identify the populations necessary for sampling and reporting. As stated earlier, a representative sample should be obtained for all of the groups shown in Table 1. If it is not possible to identify a representative sample, e.g., heritage speakers, a sample that covers the range of variability of the identified population, such as geographic representation or type of Spanish spoken, should be selected.

Students without School Learning Experience

Students who have learned Spanish without formal U.S. academic training, e.g., from use within the home, living abroad, having association with other Spanish speakers, etc. also will take the Spanish NAEP. The consensus building committees recommend reporting the achievement of these students as a group separate from both non-heritage and heritage learners who have had formal Spanish school learning experience in the U.S. The group of students *without* school learning experience also will include both non-heritage and heritage language speakers, but neither subgroup should be over sampled.

Finally, the sample of students taking the Spanish language assessment will be stratified to ensure adequate representation of relevant background variables, consistent with NAEP policy. Figure 1 illustrates stages 1 and 2 of the Spanish NAEP sampling procedure.

Figure 1: Stages 1 and 2 of the Spanish NAEP Sampling Procedure



KEY

Stage 1: LS/BQ, administered to nationally representative sample of 12th graders

Stage 2: Spanish NAEP administered to representative sample of Spanish Language Learners

Current Students: Students currently studying Spanish

Non-current Students: Students who completed their last course prior to current year

Years: Number of years of study

\+ Categories in Stage 1 of Sampling

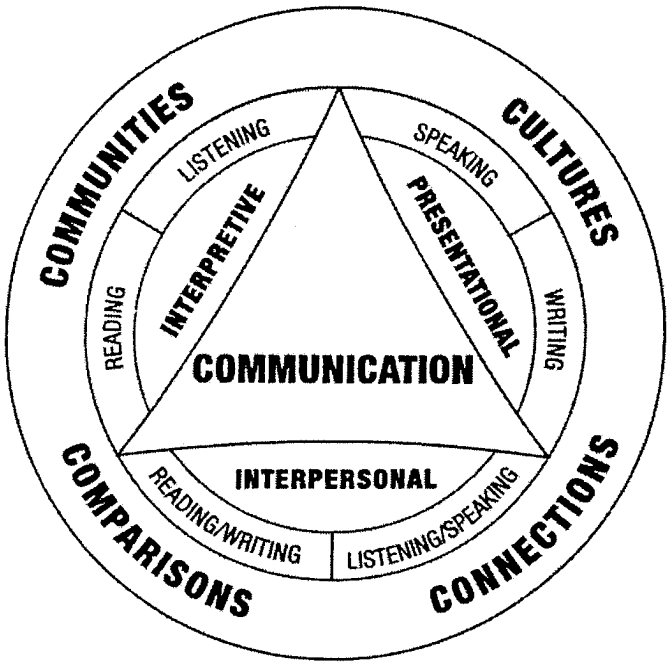
Categories in Stage 2 of Sampling

Construction of the Assessment

The Foreign Language Assessment Framework

The framework for the Foreign Language NAEP is represented by the graphic shown in Figure 2. Definitions of the terms used in Figure 2 follow.

Figure 2. The Foreign Language NAEP Assessment Framework



Evaluation Criterion

- Demonstration of Comprehension and Comprehensibility (*including the use of communication strategies and the application of cultural knowledge*)

Contexts

- Daily Life
- School
- Work

Definitions

At the recommendation of the consensus building committees (see Appendix B), the choice of the specialized terminology used in the FL NAEP framework was intentional. These terms are consistent with the widely adopted *Standards for Foreign Language Learning* and are meant to ensure that explanations are precise.

Communication

The central focus of the assessment is to measure the ability of students to communicate in a language other than English. At its heart, communication is the ability to exchange information; that is, to convey and receive messages. These messages are of many different types and may be conveyed and received in many different ways. Although language is the primary vehicle of communicative messages, being able to communicate effectively means that the individual can combine knowledge of the language system with knowledge of cultural conventions, such as norms of politeness. Knowledge of language and culture combine to make successful communication.

Modes of Communication

Although there are several ways communication can be characterized, the method here follows that of the *Standards for Foreign Language Learning*. This approach defines three modes of communication, based on the context and the purpose of the communicative interaction.

Interpersonal

The *interpersonal* mode involves two-way, interactive communication, such as conversing face-to-face or exchanging e-mail messages. It is characterized by direct communication between individuals who are in personal contact, thus allowing the participants to clarify their meaning when misunderstandings occur. In this mode, participants in the interaction use both linguistic and non-linguistic feedback from others to ascertain the extent to which their message is being successfully communicated, and can make adjustments and clarifications accordingly. Necessary to achieving successful communication in this mode are the productive language abilities of speaking and writing as well as the receptive abilities of listening and reading, and the ability to use and interpret non-verbal behavior, including gestures and body language in face-to-face interactions.

Interpretive

The *interpretive* mode relates to the one-way understanding of spoken or written language, such as listening to a broadcast or reading a magazine. It involves coming to a culturally appropriate understanding of the meaning of oral or written messages sent via print and visual images, whose original author is not present. Necessary to achieving successful communication in this mode are the receptive language abilities of listening and reading and the ability to use visual images to assist in comprehension.

Presentational

The *presentational* mode involves one-way spoken or written communication, such as giving a speech or writing a story. It involves producing spoken or written messages for an audience with whom there is no immediate personal contact. Thus, there is no possibility to clarify intended meanings when misunderstandings occur. Such messages need to reflect awareness of cultural differences in order to be presented in a manner that will enable appropriate interpretation by persons from a cultural background where the foreign language is spoken. Necessary to achieving successful communication in this mode are the productive language abilities of speaking and writing and the ability to use visual images.

Goal Areas that Provide the Basis for Communication

The *Standards for Foreign Language Learning* describe four interrelated goal areas for foreign language education that provide content, context and purpose for meaningful communication. For the FL NAEP framework, these areas provide a basis for communicative assessment tasks. All tasks will be related to at least one, if not several, of these goal areas.

Cultures

The *cultures* goal relates to the gaining of knowledge and understanding of a different culture on its own terms. Culture is understood as the perspectives, practices, and products common to a society. The philosophical perspectives of a cultural group—the meanings, attitudes, values, and ideas that form its worldview—are the basis from which practices and products are derived. Practices are the patterns of social interaction accepted by the society, such as its rules for greeting and leave-taking, how space and gestures are used in personal interactions, and how status is determined. Products may be tangible (such as books, paintings, or buildings) or intangible (such as laws or a system of education). Understanding interrelationships among perspectives, practices, and products of the culture(s) studied in the foreign language classroom provides important content and purpose for participating in communication in the foreign language. This content and purpose should be found in many FL NAEP assessment tasks. The perspectives, practices, and products of cultures within the Spanish-speaking world will be a part of the Spanish NAEP.

Connections

The *connections* goal relates to using the foreign language to connect to other academic disciplines to acquire knowledge. This goal provides a clear purpose for communication in the foreign language, whether to reinforce and further one's knowledge of other academic disciplines, or to acquire information and recognize distinctive viewpoints only available though the foreign language and its cultures. Utilizing connections as a goal area for FL NAEP assessment tasks extends the content available for assessment tasks broadly.

Comparisons

The *comparisons* goal relates to developing insight into the nature of language and culture through comparisons between the native language and culture and a foreign language and culture. Such understanding is made possible only through the study of a foreign language and culture. Making comparisons provides a clear purpose for communication in the foreign language that may be embedded in tasks on the FL NAEP. The topic of linguistic and cultural comparisons also provides additional content area for assessment tasks.

Communities

The *communities* goal relates to participating in multilingual communities at home and around the world. This goal provides a context in which the foreign language is purposefully used for communication. Within the instructional setting, the context may extend beyond the school community through communication with speakers of the foreign language via e-mail or audiotapes. Ideally, the context will be using the language for personal enjoyment and enrichment by interacting with materials and persons from the culture(s) where the language is spoken. These contexts provide situations for assessment tasks on the FL NAEP.

Evaluative Criterion

The single comprehensive criterion by which performances on FL NAEP assessment tasks will be evaluated is the demonstration of how well the student understands the foreign language (comprehension) and can be understood using the foreign language (comprehensibility). *Demonstration of comprehension and comprehensibility* includes language knowledge, and, to varying degrees depending on the assessment area, the use of communication strategies and the application of cultural knowledge. The consideration of these aspects of communicative performance is drawn from the ACTFL *Performance Guidelines for K-12 Learners* and the *Standards for Foreign Language Learning*.

Demonstration of comprehension and comprehensibility encompasses multiple aspects of the language system, such as knowledge of vocabulary, grammar, pronunciation, and spelling. Beyond these elements of the traditional foreign language course, effective communication requires control of other features including sociolinguistic elements, such as sensitivity to the social context; pragmatic elements, such as sensitivity to the communicative goals; and textual elements, including knowledge of how written and oral texts are organized. Although the importance of particular elements of this criterion vary depending on the communicative mode and assessment task, all assessment tasks on the FL NAEP will allow students to demonstrate their control of these crucial elements of the foreign language.

The use of communication strategies, such as asking for clarification or inferring the meaning of unknown words from context, enables students to engage in and maintain successful communication. Students can use communication strategies to help them understand or be understood in unfamiliar contexts or with unfamiliar vocabulary and grammar, to avoid breakdowns in communication, and to correct misunderstandings. Such strategies provide students with the ability to maintain communication despite deficiencies in language skills and cultural knowledge. Communication strategies include the ability to do the following:

- use appropriate turn-taking behavior in a conversation;
- paraphrase (say things in different ways);
- make intelligent guesses to achieve greater comprehension;
- use context to help comprehension;
- make inferences, predictions, and generalizations; and
- draw conclusions.

For the Spanish NAEP, consideration of the use of communication strategies influences the evaluation of the comprehension and comprehensibility of student performances to different degrees and in different ways. These variations depend on the communicative mode being assessed and the specific assessment task. For more information, see Chapter 6.

Finally, awareness of differences between one's own culture and the culture of the people with whom one is communicating and the ability to bridge those differences facilitates clear communication between individuals. For example, in the interpersonal mode, perspectives from the culture where the foreign language is spoken govern patterns of interaction between participants in a conversation. Good foreign language communicators are aware of how the use of formal or informal registers (that is, the variety of language appropriate to the specific social setting) and gestures differ between cultures. In the presentational mode, a sensitive communicator who is aware of cultural differences may use that knowledge to present cross-cultural background information that may facilitate communication. In the interpretive mode, students must apply cultural knowledge to interpret accurately what they read and hear.

The application of cultural knowledge differs from the *cultures* goal, which has to do with gaining knowledge and understanding. As part of the evaluative criterion, how such knowledge is used to achieve successful communication is considered. For the Spanish NAEP, consideration of the application of cultural knowledge influences the evaluation of the comprehension and comprehensibility of performances in different ways, depending on the communicative mode being assessed and the specific assessment task. For more information, see Chapter 6.

Contexts

Communication always takes place in a context, where it is used to achieve some goal or accomplish some task. Because communication occurs in a wide variety of contexts, foreign language educators are accustomed to simulating these contexts through role-plays and other methods in the classroom. Three broad context areas are appropriate for FL NAEP assessment tasks: daily life, school and work.

Daily life

Daily life refers to the use of the foreign language to participate in society outside of the school setting or to accomplish every-day social (i.e., non-academic) tasks. With the exception of some heritage language speakers and others who have had the opportunity to live or travel in a Spanish-speaking environment, students' actual use of the foreign language in daily life contexts outside the classroom may be very limited. However, through language study, students can imagine themselves in daily life contexts in which use of the foreign language may be necessary, such as shopping in a store in a foreign country, traveling abroad, or living with a host family.

School

The school context refers to the academic part of the student's life. School contexts in NAEP assessment tasks may refer to reading in the foreign language to complete an academic task or conversing about a school-related topic or theme. School as a context may reflect actual classroom experiences in the United States, or experiences in an academic environment in a foreign country (for example, as an exchange student).

Work

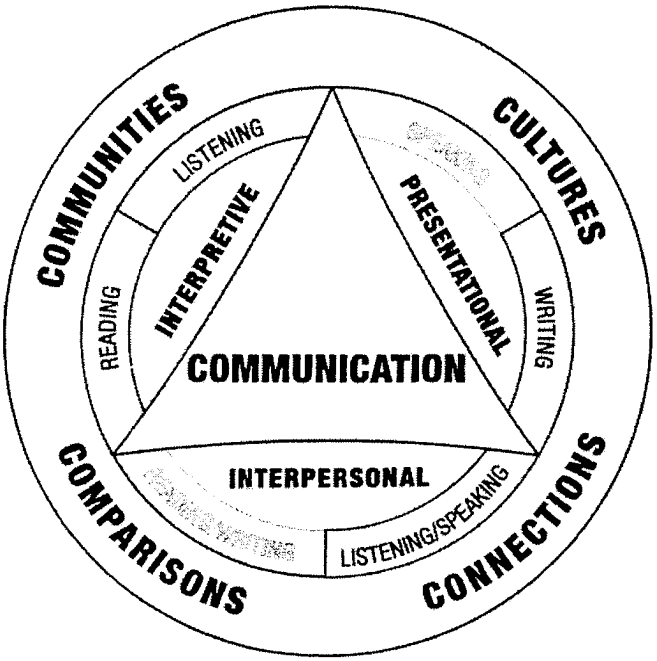
The work context refers to using the foreign language on the job, in employment situations. Work-related contexts relate to paid or volunteer experiences with which 12th graders may have familiarity.

The Spanish Assessment Framework

The framework for the Spanish NAEP is represented by the graphic shown in Figure 3. As mentioned earlier, the consensus building committees recommend that the Spanish NAEP focus on four of the six assessment areas in the general FL NAEP framework:

- listening and speaking in the interpersonal mode,
- listening in the interpretive mode,
- reading in the interpretive mode, and
- writing in the presentational mode.

Figure 3. The Spanish NAEP Assessment Framework



Evaluation Criterion

- Demonstration of Comprehension and Comprehensibility (*including the use of communication strategies and the application of cultural knowledge*)

Contexts

- Daily Life
- School
- Work

The two assessment areas not assessed by the Spanish NAEP, as shown in Figure 3, are reading and writing in the interpersonal mode and speaking in the presentational mode. There are several reasons for this. First, the consensus building committees felt that both assessment areas were used with less frequency in real-world communication situations that secondary school students might encounter than the four areas chosen. Second, reading and writing in the interpersonal mode is most effectively done through electronic correspondence. The committees felt that assessing interactive communication on-line may present logistical problems requiring a large amount of resources to solve. Finally, the committees felt that attempting to assess all six areas was potentially too costly, and it would be better to focus on assessing four areas well, particularly in light of the need to have large representative samples for the study. With the exception of its focus on four of six assessment areas, the 2003 Spanish NAEP reflects the foreign language NAEP framework described above.

Components of the Assessment

A variety of exercises (items or tasks) will be included to assess students' proficiency in the three communication modes. The exercises will assess students in the four areas identified for the Spanish NAEP—the interpersonal mode with conversation-based tasks (listening and speaking); the interpretive mode with listening-based tasks; the interpretive mode with reading-based tasks, and the presentational mode with writing-based tasks. These four assessment areas are shown in Table 2.

Table 2. Areas of Communication to be Assessed

Interpersonal Mode	Interpretive Mode	Presentational Mode
A. listening/speaking (conversation-) based tasks	B. listening-based tasks C. reading-based tasks	D. writing-based tasks

All students (100% of the sample) will take tasks in two of the three assessment areas B, C, and D. Multiple blocks (or sets) of items will be created within each area. Combinations of blocks must be created to allow for construction of scales within each of these three areas (B, C, and D) and to allow for estimation of the relationships in performance among the three areas.

The examination time (length of test sitting) for each student is 50 minutes. Approximately equal testing time should be allocated to each of the three areas, B, C, and D.

Because of the costs of administering and scoring the conversation-based tasks, only a subset (20%) of the sample of students taking the Spanish assessment will be administered tasks in assessment area A. Possible combinations for these students include: A-B-C, A-B-D, and A-C-D. Students being assessed with tasks in assessment area A have an additional 20-25 minutes of testing time.

Construction of the Spanish Language Assessment Exercise Pool

Distribution of Exercise and Response Formats

Specifications for distribution of exercise and response formats across the four areas of the Spanish assessment are shown in Table 3.

Table 3. Specifications for Exercise and Response Formats

Assessment Area	Exercise Format	Student Response/Scoring
A. Interpersonal: Listening/Speaking (Conversation) based Tasks	One-on-one conversation (test administrator and student)	Video taped Conversation in Spanish (Spanish to Spanish) Holistically scored on one scale having multiple dimensions
B. Interpretive: Listening-based Tasks	CD ROM (individual student) or video (large group) Listening stimuli in Spanish Questions in English	Multiple-choice format (options in English) Short answer format (student responds in English) Computer-based or paper/pencil
C. Interpretive: Reading-based Tasks	Texts in Spanish Presented in paper format Questions in English	Multiple-choice format Short answer format (student responds in English) Paper and pencil
D. Presentational: Writing-based Tasks	Directions for writing tasks in English Presented in paper format	Short answer (student responds in Spanish) Extended answer (student responds in Spanish) Paper and pencil

To ensure that students understand what is expected of them, all test directions will be in English. As shown in the table above, for the conversation-based exercises, the interaction will be entirely in Spanish (though directions will be in English). The conversations will be captured on videotape. The stimuli for exercises in the interpretive mode will be in Spanish, but the questions and responses (multiple-choice or short-answer format) will be in English, unless the task is to choose an appropriate Spanish text from among several options. The stimuli for the listening-based exercises will be presented on CD-ROM to individual students or by video to groups of students. The responses will be either computer-based or pencil and paper responses. Both the stimuli and responses for the reading-based exercises will be in paper format. The directions for the writing-based exercises in the presentational mode will be in English and presented in paper format; the written responses will be given in Spanish, also in paper format.

Students administered the Spanish NAEP will need to have a command of English in order to demonstrate their achievement in Spanish on the test. Standard NAEP accommodations will be made for students with disabilities and for limited English proficient students, no matter what their language background. The FL NAEP is not intended as a test of bilingualism or of Spanish as a first or native language. It is intended to enable students in an English-language school setting to demonstrate achievement in a language other than English; in the case of the Spanish NAEP, the language is Spanish.

Distribution of Exercises Across the Assessment

The consensus building committees recommend that, for the interpretive areas (reading and listening) and the presentational area (writing), the Spanish NAEP be targeted to assess different levels of language proficiency.[2] For these communication areas, a minimum of two levels of the assessment (i.e., two sets of test booklets composed of randomly selected blocks of exercises from the three areas) that vary in difficulty should be constructed to provide appropriate assessment of the wide range of language proficiency expected among 12th grade students.

The different levels of the assessment should be designed to cover content specified by the Framework document. Therefore, a matrix sampling of test content, similar to that done in other NAEP assessments, should be combined with targeted testing. This will require at least two sets of test booklets that vary in test difficulty. Booklets must be constructed so that 1) valid scales can be constructed within each of the three assessment areas, 2) a vertical scaling can be conducted across the different difficulty levels of test forms, and 3) correlations among the areas can be estimated. Scaling and vertical linking procedures must be used to link scores from different test booklets and difficulty levels of the assessment to a single score scale for each area. These procedures should include overlapping test booklets that are at similar difficulty levels as well as overlap across test booklets at different difficulty levels.

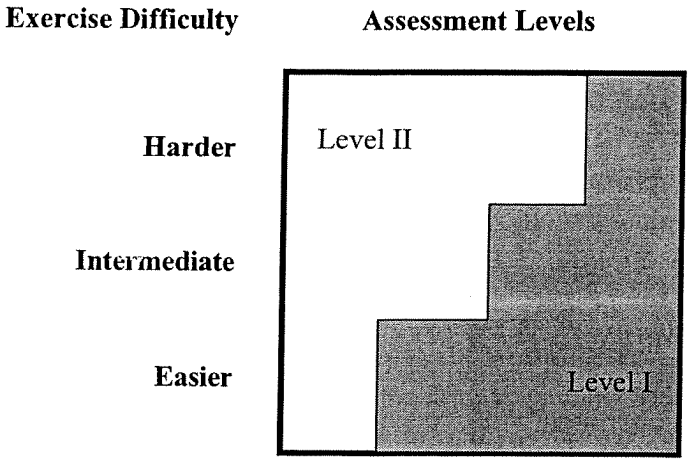
Students will be routed into test booklets of appropriate difficulty through multiple sources of information—self reporting of proficiency, self reporting of years of study (cross-validated by checking of school records where possible), report of foreign language experiences, teacher or counselor recommendation, and a brief questionnaire on Spanish vocabulary knowledge. We anticipate that students who have taken fewer years of academic study or who have had some limited exposure to Spanish learning outside of school will take easier test booklets. Students who have taken more years of study, have engaged in foreign study or travel, or have been educated in Spanish abroad will take more difficult test booklets.

Exercises will be designed to provide appropriate assessment of students over a wide range of achievement, including those at both the low and high ends of the proficiency spectrum. At the low end, the exercises should provide appropriate assessment of students who are just beginning to study Spanish (1 year), those who are currently in their second year of study, and those who completed one or two years of study in the past; i.e., non-current language students. At the high end, the exercises should provide tasks of sufficient richness and depth to allow high achieving students and those who have pursued in-depth study to demonstrate the extent of their achievement.

[2] There will be only one level of the interpersonal assessment.

Figure 4 illustrates how, if only two levels of the assessment are administered, exercises may be distributed across them, assuming three levels of exercise difficulty. Both levels of the assessment contain exercises at each difficulty level. However, Level I contains a higher proportion of easier exercises, whereas Level II contains a higher proportion of the more difficult ones.

Figure 4. Distribution of Exercises Across Two Levels of the Assessment



There are several considerations for deciding the number of difficulty levels of the assessment that will be needed: 1) the classification accuracy and consistency that are obtained from the routing procedure, 2) the distribution of language ability in the population, 3) the amount of time available for the assessment, and 4) the number of items that can be administered in that time; i.e., test length. The Technical Advisory Panel recommends that there be two levels of the assessment per assessment area, but if, for accurate assessment, more levels are needed, or if testing time constraints limit the number of levels, more levels or a single level can be considered for each assessment area. The decision about number of levels of the assessment should be made *after* the field test in order to see the range of ability represented in the population. Data from the field test will be needed to inform decisions about the number of assessment levels, as well as how to assign items of differing difficulty to test booklets within the levels.

There also must be some overlap of exercises between different levels of test booklets. The vertical linking requires that the quality of the field test data be such that it can be used to make assignments of items into booklets within the assessment levels. Examinees will be assigned to test booklets based on the routing process described above.

The testing contractor must develop procedures for ensuring that exercise blocks in the different levels of the assessment, sampling of examinees, screening examinees into assessment levels, and the matrix sampling work together to yield accurate information about the Spanish language proficiency of 12^{th} grade students in the U.S. who have studied Spanish in a variety of ways and for different lengths of time.

Although targeted testing (i.e., testing by level) is recommended only for the interpretive and presentational modes, the testing contractor also must assemble interpersonal assessment booklets (at a single level) that will have an overlap of exercises and will allow creation of an interpersonal communication scale. The administration of the interpersonal assessment block and the other assessment blocks should be such that correlations between the interpersonal mode and the other communication modes can be estimated.

Spanish Vocabulary Questionnaire

Examinees will be assigned into levels of the assessment based on the routing process described above, including a short vocabulary questionnaire. The Technical Advisory Panel recommends that a short questionnaire on Spanish vocabulary knowledge be considered for use as part of the routing procedures, as recent research shows high correlations between vocabulary knowledge and the skills of listening, speaking, reading, and writing.

For example, research shows that successful communication is heavily dependent upon a knowledge of individual word meanings. The widely recognized relationship between vocabulary and comprehension attests to the crucial role word knowledge plays in text understanding, both in a first language (e.g., Anderson & Freebody, 1983; Carroll, 1971; Davis, 1968; Beck, McKeown, & Omanson, 1987; Mezynski, 1983) and a second language (e.g., Carrell, 1988; Alderson & Urquhart, 1985; Laufer, 1991; Ulijin & Strother, 1990). In addition, research repeatedly shows that vocabulary knowledge correlates more highly with reading comprehension than other linguistic and non-linguistic factors, such as knowledge of grammar and word forms (e.g., Qian, 1998; Koda, 1989; Laufer & Sim, 1985), reading strategies (Haynes & Baker, 1993), and inference skills (e.g., Anderson & Freebody, 1981). In summary, vocabulary knowledge is viewed as an integral aspect of overall language ability.

All students who will take the Spanish NAEP will have taken the vocabulary questionnaire. Because the use of a routing process to assign students to test levels is new to NAEP, and because of the importance of accurate assignment of students, the Technical Advisory Panel recommends small-scale piloting of the routing procedures, as well as field testing them when field testing the cognitive items, in order to ascertain the accuracy of the procedures. Measures will be taken to ensure student and school anonymity in collecting these data on-line.

Review Process and Criteria

All exercises will be subjected to multiple review steps. Exercises for assessment areas B, C, and D, will be reviewed in the medium in which they will be administered, i.e., reviewers will see electronic versions of those exercises that will be administered electronically and paper versions of those that will be administered on paper. Exercises developed for the interpersonal area will be reviewed in paper form.

Spanish Expert Review

There will be separate exercise pools for the different assessment areas (A,B,C,D). To ensure the development of exercises (items) that adequately represent the content domain and exhibit proper psychometric characteristics, as well as to construct exercise pools that will adequately measure the processes, skills, and knowledge described in the achievement levels, review by Spanish educators who are competent in test construction and knowledgeable in relevant Spanish language content must be incorporated at several points during the assessment development process. Therefore, the development, field testing, and selection of exercises should be monitored by an item development panel consisting of members of the NAEP Foreign Language Consensus Planning and Steering Committees, teachers, Spanish educators, and other subject area experts. A minimum of 20 percent of the membership of the item development panel will be from the FL NAEP framework and specifications consensus committees, as specified by National Assessment Governing Board policy.

After the exercises have been developed, the panel will review the pools and judge the exercises for congruence with the framework and specifications documents. The exercises will be judged on criteria such as appropriateness and authenticity of the Spanish language used, technical accuracy, content validity, variety of formats, and the mode of communication that references the exercises to the assessment dimension they purport to measure. Exercises also will be reviewed for age appropriateness, their ability to engage the students and the authenticity of the tasks. Finally, reviewers will ensure that the available pools are balanced, that they are representative of the content and skills described in the achievement level definitions, and that they incorporate sufficient exercises at the various achievement levels.

Exercises will be reviewed again after field test administration, as part of the process of selecting those that will appear in the operational assessment. Any exercises that statistical evaluation reveals to be technically flawed will be eliminated.

Language Simplification Review

All test directions and all test exercises that are written in English will be subjected to review by a panel of experts recruited from the Limited English Proficiency and Student Disabilities communities. In an effort to ensure that an inordinate amount of English language knowledge is not a factor in accessing assessment tasks, the panel will review all directions and exercises written in English using guidelines for language simplification consistent with other NAEP assessment development activities.

Bias Review

All exercises will be screened for lack of ethnic, cultural, and gender sensitivity, and will be subjected to Differential Item Functioning (DIF) analyses following accepted psychometric practices. The field test administration samples will be selected to be as representative of the NAEP operational sample as possible.

If, after close scrutiny, an exercise appears to be a valid measure of Spanish language content and communication ability, and if no plausible explanation for a differential performance is apparent, the exercise will be retained. As mandated by law, the National Assessment Governing Board has final review authority of all cognitive items. The Board reviews all cognitive items prior to field-testing and, subsequently, prior to the operational assessment.

Administration Specifications

Selection of Administrators

The administrators of the Spanish Language Assessment should be knowledgeable about and good communicators in the Spanish language. If possible, they should be current or former practitioners in the discipline. The interpersonal exercises will require students to participate in one-on-one task-based conversations in Spanish with the test administrator. It is essential that the administrators who conduct this portion of the assessment be fluent speakers of Spanish and have experience speaking with Spanish language learners. They must have experience administering oral proficiency interviews. All other persons hired to administer the Spanish Language Assessment should meet the following minimum qualifications: background and training in Spanish and experience teaching students at the high school level. Experience working with heritage speakers is desirable. In addition, it will be useful for administrators to have experience with video equipment, both for administering the interpretive listening stimuli in Spanish to groups of students and for videotaping the interpersonal exercises.

Training of Administrators

Extensive training will be provided to ensure that administrators are appropriately prepared to administer the Spanish language assessment. The training materials should cover topics such as room arrangements for students; preparation for testing, including set-up and employment of technology required to administer the assessment; administration procedures, including preparation of students for videotaping of responses; and guidelines for the amount and type of interaction that can be allowed between examinees and administrators.

Additional training will be provided to those administrators who will conduct the one-on-one conversations with students for the interpersonal block. Administrators of the interpersonal block should view exemplary videos from the field test so that they can gain an understanding of the interpersonal assessment process and the target behaviors desired. These administrators also should have the opportunity to role play the assessment experience with their peers and then practice administering the assessment to students with coaching to improve their effectiveness. Since student responses to the interpersonal exercises are to be videotaped, the administrators should receive instruction on how to carry out those functions.

Training of the administrators of the interpersonal exercises is critical to ensure standardized conditions to facilitate scaling of the interpersonal component of the assessment. The testing contractor will need to conduct inter-administration reliability checks immediately after training and again after the field test.

Resources

Resources will be needed for the administration of the Spanish Language Assessment. The interpretive listening exercises will require videotape clips to which students are asked to respond. A VCR and TV monitor will be needed for showing and viewing the videotapes. If listening stimuli are administered by CDROM to individual students, and the questions and responses are computer based, personal computers will be required.

For the interpersonal exercises, a physical setting that is free from disturbance and has good lighting will be required so the students can be videotaped. Videotaping capacity (camcorders, microphones) for recording the conversations between the students and administrators during the interpersonal exercises also will be required.

Ethical Considerations

The Spanish Language Assessment exercises will be sensitive to the privacy of students and not ask them to reveal sensitive personal information or experiences. Further, the assessment exercises will not ask students about privileged home information or psychological information. In addition, the ethnic and cultural sensibilities of the students will be respected by choosing only tasks and topics for exercises that could not be construed as insulting or biased against any ethnic, racial, religious, geographic, or cultural group. Finally, the assessment will provide guidelines for the protection of any videotapes that capture student responses. A letter of permission will be sent to parents of students participating in the assessment that describes the uses to which the videotapes will be put and explains that no videotapes will be able to be traced by any person, including scorers, to individual students or schools. Such videos will be stored in a secure facility and will not be shown except under secure conditions for the training of raters.

Scoring and Reporting Specifications

General Criteria

The assessment should be designed to facilitate reporting using NAEP achievement levels, quantitative scales, and narrative descriptions of typical exercise solutions by students. Estimates of proficiency distributions for the population as a whole, and for subgroups of the population, will be constructed using statistical procedures that appropriately account for the imprecision in the measurement of individuals, in accordance with standard NAEP procedures. A detailed discussion of subgroups that should be reported follows.

Spanish Communication Score and Subscales

Sufficient exercises will be included in the assessment to assure the precision necessary to report results on four communication subscales: interpersonal mode (conversation), interpretive mode (listening), interpretive mode (reading), and presentational mode (writing). Booklets will be assembled so that valid scales can be constructed within each of the four areas, and correlations among the areas can be estimated.

Reporting these scores separately does not preclude the reporting of a composite Spanish communication score. However, attention needs to be paid to subgroup differences across performance dimensions (communication modes) and the appropriateness of forming a composite. Correlations across scales should be examined among subgroups. If correlations are not really strong among scales, then the scales should not be included in a composite scale. If a composite score is to be created, it should be an appropriately weighted combination of the four scale scores and labeled to convey the idea that it is a composite of separate scales.

Evaluation Procedures

Performances on assessment tasks will be evaluated on the demonstration of how well the student understands (comprehension) and can be understood (comprehensibility). As discussed earlier, in addition to language knowledge, this criterion includes considerations of the use of communication strategies and the application of cultural knowledge, to varying degrees and in different ways depending on the communicative mode and specific assessment task.

Scoring of Open-ended Exercises

Scoring student responses to open-ended questions will be consistent with previous practices in NAEP. Scoring guides (rubrics) will be used for scoring the constructed-response exercises (short and extended response) and for the conversation exercises.

Constructed Response Exercises

Some short constructed-response exercises that appear on the assessment in the interpretive and presentational areas may call for responses that are to be dichotomously scored right/wrong (i.e., credit/no credit scoring). Other constructed response exercises included in these areas will be designed so that varying amounts of credit can be awarded depending upon the quality of the student's response (e.g., partial credit scoring using a scoring guide that has several score points, each defining a level or category of performance). Scoring guides for short constructed-response exercises designed to allow partial credit should consist of three levels, with specific criteria provided for responses at each level. Extended constructed-response exercises should be scored using a four or five point scale in order to obtain more in-depth information from these longer, more complex student responses.

Interpersonal Communication Exercises

Each interpersonal conversation will be scored holistically on a single scale spanning multiple dimensions. A scoring guide will be developed for *each task* in the interpersonal exercise using a four to six-point scale, with specific criteria provided for responses at each level, and a single score will be given for each task. The individual task scores will be aggregated to reach a single final score for the interpersonal mode exercise.

Scoring guides should be created for the constructed-response exercises (short and extended response) and for the interpersonal conversation exercises as the exercises are being developed. After the exercises have been field tested, the scoring guides should be refined based on actual student responses. Points should be assigned within the scoring guides for complex items to reflect the quality of the responses, and each point on the scoring guide should contain specific criteria for the desired response at each level. The scoring guides must adhere to the requirements for content as defined in the assessment framework and specifications, and there must be a tight match between the demands of the exercise task and the criteria contained in the scoring guide.

In addition, the assessment exercises must be designed so that the components to be scored are quite evident. Requirements of each question should communicate clearly to students what is being asked and how, in general, their responses will be evaluated. Criteria should be distinctly defined so that raters will understand clearly how to evaluate the student responses. Finally, each score point in the scoring guide should be sufficiently differentiated to allow raters to apply the various points on the scale consistently. An appropriate IRT model for polytomous exercise response data should be used in scoring the constructed-response and conversational exercises to accommodate a range of possible scores, e.g., 0 to 5, rather than just assigning a right or wrong score.

Score Reporting

The primary means of reporting NAEP results is by achievement levels—*Basic*, *Proficient*, and *Advanced*. The data should be presented in terms of the percentage of students at or above the *Basic*, *Proficient*, and *Advanced* achievement levels for each area of the assessment. Preliminary Achievement Level descriptions are presented in a separate section later in this document.

The preliminary recommendations for NAEP achievement levels are specific to the different modes of communication presented on the assessment and have been informed by the *Performance Guidelines for K–12 Learners*. Unlike other subjects areas assessed by NAEP, performance standards as defined by the ACTFL *Proficiency Guidelines* for foreign language skills are already widely disseminated and used as a standard in foreign language education in the United States. Of more importance to NAEP, however, are the ACTFL *Performance Guidelines for K-12 Learners*, which appeared in 1998. Recognizing that the original *Proficiency Guidelines* were developed for assessing adult language abilities, the ACTFL *Performance Guidelines for K-12 Learners* focus on second language use by students in elementary, middle and high school foreign language programs. The section on Preliminary Achievement levels later in this document provides more description of the relationship between the NAEP achievement levels and the *Performance Guidelines for K–12 Learners* and provides preliminary definitions for each level in each assessment area.

No single method of reporting of NAEP results has been optimal for all purposes. Consequently, multiple methods of reporting results should be investigated. In particular, (scale) scores and/or percentages should be supplemented with narrative descriptions for individual illustrative items or clusters of items. As mentioned above, some short constructed-response exercises appearing on the assessment will call for responses that are to be dichotomously scored. Illustrative exercises of this type should have results reported in terms of percent correct. Similarly, results

for illustrative multiple-choice items should be reported in terms of percent correct. Other constructed response exercises will be designed so that varying amounts of credit can be awarded depending upon the quality of the student's response. Illustrative exercises of this type should have results reported in terms of the percentage of student responses falling into each performance category. A description of the scoring guide for each illustrative exercise should be presented, as well as examples of student responses illustrating each performance category included in the scoring guide.

Reporting Subgroup Performance

Results for representative samples of students who have studied Spanish for varying amounts of time, heritage language learners, and students who have learned Spanish without formal U.S. academic training should be reported. Table 1, repeated here, shows the groups of students for whom results should be reported. Results also will be reported for the standard NAEP subgroups.

Table 1. Categories of Students

1. Non-Heritage Language Students With U.S. School Experience Learning Spanish, Currently Studying Spanish in 12th Grade	
1A	5 or more years of continuous study (i.e., study was begun in 8th grade or earlier)
1B	4 years of continuous study (i.e., study was begun in 9th grade)
1C	3 years of continuous study (i.e., study was begun in 10th grade)
1D	2 years of continuous study (i.e., study was begun in 11th grade)
1E	1 year of study (i.e., study was begun in 12th grade)
2. Non-Heritage Language Students With U.S. School Experience Learning Spanish, Who Have Completed Their Last Spanish Course Prior to Current Year	
2A	3 or more years of study
2B	2 years of study
2C	1 year of study
3. Spanish Heritage Language Speakers with U.S. School Spanish Learning Experience	
4. Spanish Speakers (Heritage and Non-Heritage) without U.S. School Spanish Learning Experience	

When reporting results by years of study and for demographic subgroups, procedures should be used that are consistent with current NAEP practice. Reporting of subgroup performance may be both by subscales and a weighted composite of subscales. If such a combination of scale scores is used, reports of overall score distributions should make clear that they are a composite of separate subscales. Details of score reporting are presented in a separate report entitled *The Foreign Language NAEP Reporting Strategies Document.*

Recommendations for Special Studies

As discussed in the Framework document, foreign language education in the United States is a patchwork of different program models with various entry and exit points. For the first foreign language NAEP, the National Assessment Governing Board mandated an assessment in Spanish for secondary school learners. Recognizing the limitations of the main assessment in terms of the age of the learners and the language chosen, the Board asked the consensus building committees to propose options for a possible small-scale special study that would gather additional data on students' foreign language achievement beyond that collected in the main assessment.

Because there are so many options for studying achievement in other languages and at other levels of study (elementary and middle school), the consensus building committees carefully weighed various alternatives. In the end, three important studies were identified. Recognizing that only one may be chosen, the FL NAEP consensus building committees prioritized the studies in the order shown below.

1. The Achievement of Early Language Learners in Japanese

NAEP has traditionally conducted benchmark testing of our nation's students at Grades 4, 8, and 12. For the foreign language assessment, the main study will be conducted solely at the twelfth grade level. However, more and more school systems are implementing K–12 programs and moving toward the vision of well-articulated, long sequences of language study as set forth by the *Standards for Foreign Language Learning.* Thus, the decision to test at just the 12th-grade level inaccurately reflects the reality of foreign language education in the country. By conducting a small-scale study at the 4th-grade level, our nation's educational policymakers will have information on the first stages of achievement for students who begin the study of a foreign language in elementary school.

Professionals who have worked at the elementary level know that students beginning the study of a foreign language in kindergarten or first grade can make enormous strides by Grade 4. Specifically, students who start early generally exhibit strong oral language skills, which are crucial in the interpersonal mode. Conversely, achievement in the interpersonal mode is often the weakest area for high school students.

A small scale study conducted at the 4th-grade level would provide data for examining the specific advantages of an early start to learning foreign language. The study could focus on the question of which program model (Foreign Language(s) in the Elementary School, Partial Immersion, or Total Immersion—see Framework document) is most effective in developing second language proficiency. As more and more school systems begin to implement programs at the elementary level, the information gained from such a study would be valuable for providing guidance and direction for future programs.

Because the main NAEP is in Spanish, choosing a language other than Spanish for the small-scale study has certain merits. It is essential for policy makers to understand that there are tremendous benefits to be gained from a study of any foreign language. Because of the contrast to Spanish, Japanese offers several advantages for the small scale study. Assessing achievement in Japanese could help demonstrate what younger learners are capable of achieving in a language

generally perceived as very difficult to learn. In particular, because Japanese has a different writing system, a Japanese study could shed light on the effects of early foreign language learning on the acquisition of a symbol system so different from English. In addition, Japanese culture is very different from U.S. culture. The role of culture in learning languages among young learners also could be investigated.

2. The Achievement of 12th Graders in Japanese

The basic NAEP question is—what can our students do in foreign language by the time they graduate from school? A small-scale study of 12th-grade students learning Japanese will strengthen the FL NAEP. Achievement in foreign language learning may vary by the language studied. This variation only will be evidenced if two languages are part of the FL NAEP. Assessing the achievement of 12th-grade students not only gives us the outcomes of school language learning, but also provides us information about the entry level for those who may continue language study at the post-secondary level.

The assessment of the achievement of students who have studied Japanese would also offer insights about the level of accomplishment in a language that, unlike Spanish, has few similarities with English. Likewise, opportunities for exposure to Japanese outside of the language classroom are much more limited than for Spanish. In these respects, Japanese has more in common with other languages taught in the United States (such as Russian, Chinese, and Arabic) than does Spanish. The 12th-grade Japanese assessment could be developed using the FL NAEP framework and thus be as comparable as possible to the Spanish NAEP. Further, a fair proportion of heritage speakers may be represented in the sample. As Japanese has one of the largest enrollments of any less commonly taught languages, there is a reasonable number of high school programs available for testing. Finally, unless a language other than Spanish is also included in the FL NAEP, there is the possibility that NAEP will have little influence in the language community at large.

3. The Achievement of Early Language Learners in Spanish

A small scale study conducted at the 4th-grade level in Spanish, similar to the one described in priority 1, is the third priority of the consensus building committees. While recognizing that Spanish language education does not equal foreign language education in the United States, a small scale study in Spanish has some merits. By conducting the small scale study in Spanish, it may be possible to compare, to some degree, results of 12th-grade students with those of 4th-grade students. In addition, the majority of immersion programs in this country are in Spanish, so sampling students would not be a problem. The information gained from the small-scale study for people involved in these programs would be valuable as they continue to refine and articulate the curricula of these programs.

CHAPTER 2. EXERCISE SPECIFICATIONS FOR THE SPANISH LANGUAGE ASSESSMENT

General Characteristics of Exercise Development

All exercises should test only the content described in the specifications for each communication mode. Exercises should focus on key concepts, principles, and skills appropriate for measuring communication proficiency in the various modes and should avoid testing obscure or esoteric material.

Because the emphasis for the FL NAEP is on performance; i.e., what students *can do* with their communicative language skills, assessment exercises for the FL NAEP should require demonstration of the accomplishment of communicative tasks. Exercises should *not* assess what students simply know about the foreign language (e.g., knowledge of grammar rules), or their abilities in isolated skills, such as pronunciation.

Authenticity

To the greatest degree possible, exercises or tasks for the Spanish NAEP should be authentic. Although there are various interpretations of this term, for the Spanish NAEP it means that Spanish language used in the NAEP should be real language as used by real people in Spanish-speaking communities and cultures, and not language created for instructional purposes of non-native speakers. Thus, for example, reading-based tasks assessing the interpretive mode should include texts written for real-world communicative purposes that are modified only to the extent necessary when taken out of the original context. In other words, writers may modify intact excerpts from longer texts by adding information from the longer text necessary to make the excerpt comprehensible to students who do not have knowledge of the original text. Visuals that accompany original documents must be maintained. Authenticity also means that communicative purposes of the assessment tasks should represent those of real-world language use to the greatest degree possible.

Engagement

Exercises on the Spanish NAEP should contain content and topics appropriate to the interests and experiences of 12^{th} graders. In addition, each assessment task should have a communicative purpose that is sufficiently engaging to the interests and experiences of students to ensure motivation for completing the task. The presentation of each task (for example, the use of visuals to accompany directions) also should serve to motivate and engage students.

Language Variety of Assessment Stimuli

Although there are many different varieties of Spanish and local dialects that are very distinctive, the type of Spanish used for Spanish NAEP assessment tasks should be the type most widely understood in the Spanish-speaking world. Language that is limited in usage to a certain local area or region should be avoided.

As discussed earlier, exercises should be designed to provide appropriate assessment of students over a wide range of achievement, including those at both the low and high ends of the proficiency spectrum. At the low end, exercises should be developed to provide appropriate assessment of students who are just beginning to study Spanish (Level 1), those who currently are in their second year (Level 2) of study, and those who completed two years of study some time past, i.e., non-current language students. At the high end, the exercises must provide tasks of sufficient richness and depth that high achieving students and those who have pursued in-depth study can demonstrate the extent of their accomplishments.

In addition, exercises must be developed in such a way as to ensure that the item pool in each of the assessment areas is congruent with the framework and corresponds to the following achievement level descriptions.

Preliminary Achievement Level Descriptions

Achievement levels describe what students should know and be able to do. The levels presented below define appropriate expectations for the communication proficiency in Spanish of twelfth grade students as measured by NAEP.

Three achievement levels—*Basic*, *Proficient*, and *Advanced*—have been established by the National Assessment Governing Board for NAEP assessments, as shown in Table 4.

Table 4. Achievement Level Policy Definitions

Basic	This level denotes partial mastery of prerequisite knowledge and skills that are fundamental for proficient work at each grade.
Proficient	This level represents solid academic performance for each grade assessed. Students reaching this level have demonstrated competency over challenging subject matter, including subject-matter knowledge, application of such knowledge to real-world situations, and analytical skills appropriate to the subject matter.
Advanced	This level signifies superior performance.

Performance that fails to reach the *Basic* level is categorized as below *Basic*.

As mentioned earlier, performance standards as defined by the ACTFL *Proficiency Guidelines* for foreign language skills are already widely disseminated and used as a standard in foreign language education in the United States. Describing general proficiency in each of the four skills (speaking, listening, reading, and writing), the ACTFL *Proficiency Guidelines* identify four main

performance levels: "Novice," "Intermediate," "Advanced" and "Superior." Each of the three lower levels is divided into three sub-levels, designated "low," "mid" and "high." Each sublevel is accompanied by a description of the performance represented by the level.

More important to the FL NAEP, however, are the ACTFL *Performance Guidelines for K-12 Learners*, which appeared in 1998. Recognizing that the original *Guidelines* were developed for assessing adult language abilities, the ACTFL *Performance Guidelines for K-12 Learners* focus on second language use by students in elementary, middle and high school foreign language programs. The *Performance Guidelines* are more aligned with the *Standards for Foreign Language Learning* than are the ACTFL *Proficiency Guidelines*. Organized by communicative mode (interpersonal, interpretive, and presentational) rather than language skill, their language descriptors are organized into six domains: comprehensibility, comprehension, language control, vocabulary, cultural awareness, and communication strategies. Unlike the *Proficiency Guidelines*, the *Performance Guidelines for K-12 Learners* are only divided into three main levels (novice, intermediate, and pre-advanced), without further subdivisions.

Although the preliminary NAEP achievement levels are specific to the modes of communication presented on the assessment (as opposed to the more generic descriptors of the ACTFL *Guidelines*), the preliminary recommendations for these achievement level descriptions have been informed by the *Performance Guidelines for K-12 Learners* in particular. More specifically, the descriptors for the intermediate learner have influenced those for the NAEP "proficient" level. It must be kept in mind that the intermediate learner descriptors are assumed to describe the performance of students completing either a 9-year articulated program in grades K-8, or a 6-year articulated program in grades 7-12. Although the number of students in such language programs remains small, only such students have had the opportunity to gain competency over challenging foreign language subject matter and to develop independent communicative skills in the foreign language. Students completing the most common 2-year sequence of study are assumed most likely to end up in the "novice learner range," as described by the *Performance Guidelines for K-12 Learners*. This range has informed the preliminary recommendations for the NAEP "basic" achievement level, whereas the descriptions for the "pre-advanced learner" have informed the preliminary recommendations for the "advanced" NAEP level. In this way, the ACTFL *Performance Guidelines for K-12 Learners* have been related to the NAEP performance levels.

Separability Between Communicative Modes

Although the assessment of how well students can communicate in a foreign language is the primary goal of the FL NAEP, it is recognized in practice that communication is a complex activity and that skills can develop disparately in the different modes. For example, pre-literate individuals may be able to participate in interpersonal conversations, but unable to create or interpret written messages. In academic programs, development of communicative competencies may be a function of the varying emphases of different programs. For example, traditional language instruction has emphasized reading skills, whereas today many programs emphasize speaking. For some students, skills across the communicative modes may be highly correlated; for others, the relationships may be weaker.

Because of this differentiation, the preliminary recommendations for the Spanish NAEP achievement levels are separated by the three communicative modes, with the interpretive mode being further separated for reading-based tasks and listening-based tasks. Nevertheless, although results on the Spanish NAEP will be reported in terms of NAEP achievement levels for the four assessment areas separately, this does not preclude them being combined to provide a single composite communication score.

The consensus building committees recommend the following preliminary achievement level descriptions for the 2003 NAEP Spanish Language Assessment. It is assumed that every higher performance level incorporates and builds upon the preceding levels.

The primary audience for these preliminary foreign language achievement level descriptions is the team of NAEP item writers. Since NAEP has never assessed foreign language, the preliminary descriptions were developed separately for each assessment area to give maximum guidance to item writers. The resulting item pool should validly and reliably measure the framework content and skills at the *Basic*, *Proficient*, and *Advanced* levels. These separate descriptions for item development purposes do not preclude a set of more global descriptions for *Basic*, *Proficient*, and *Advanced*. As has been the case with more recent NAEP level-setting procedures, these detailed descriptions will need to be modified and condensed prior to actual achievement level setting and reporting to the public. This revision step should occur after the initial round of item development and field testing, to allow for an examination of the full item pool.

Preliminary Recommendations for Achievement Levels

Table 5. Interpretive Mode: Reading-Based Tasks

ACHIEVEMENT LEVELS		
Basic	**Proficient**	**Advanced**
Students at this level demonstrate partial mastery of the knowledge and skills needed to comprehend ideas in written foreign language texts.	*Students at this level demonstrate a solid mastery of the knowledge and skills needed to comprehend ideas in written foreign language texts.*	*Students at this level demonstrate superior performance in interpreting the ideas in written foreign language texts.*
Students functioning at the *basic* level should be able to:	Students functioning at the *proficient* level should be able to:	Students functioning at the *advanced* level should be able to:
Identify basic topic of the text	Identify main ideas and some specific details on familiar topics	Describe the more abstract themes and ideas of the overall text
Show some evidence of making inference and predicting	Make inferences and logical predictions, as well as identify the author's purpose	Evaluate effectiveness of how the text carries out the author's purpose
Draw some conclusions and occasionally support them with information from the text	Draw conclusions and support them with information from the text	Analyze and support their analyses with examples from the text
Use own culture to comprehend the text	Use knowledge of both own culture and cultures of the Hispanic world to interpret text	Use knowledge of both own culture and cultures of the Hispanic world to develop perspectives on the text
Identify some products and practices of the cultures of the Hispanic world	Identify the interrelationships among the perspectives, practices, and products of cultures of the Hispanic world	Explain interrelationships among the perspectives, practices, and products of cultures of the Hispanic world
Recognize some differences in the products and practices of the cultures of the Hispanic world and their own	Recognize differences and similarities in the perspectives of cultures of the Hispanic world and their own	Compare and contrast differences and similarities in the perspectives of the cultures of the Hispanic world and of their own

Table 6. Interpretive Mode: Listening-Based Tasks

ACHIEVEMENT LEVELS		
Basic	**Proficient**	**Advanced**
Students at this level demonstrate partial mastery of the knowledge and skills needed to comprehend ideas in spoken foreign language passages.	*Students at this level demonstrate a solid mastery of the knowledge and skills needed to comprehend ideas in spoken foreign language passages.*	*Students at this level demonstrate superior performance in interpreting the ideas in spoken foreign language passages.*
Students functioning at the *basic* level should be able to:	Students functioning at the *proficient* level should be able to:	Students functioning at the *advanced* level should be able to:
Understand short, simple conversations and narratives, as well as recorded material in familiar contexts	Understand extended conversations and narratives as well as recorded material in familiar contexts	Understand extended conversations and narratives as well as recorded material in a variety of contexts
Identify main ideas on familiar topics found in TV, radio, video, or live and computer-generated presentations	Identify main ideas and some significant details on familiar topics found in TV, radio, video, or live and computer-generated presentations	Identify main ideas and some specific information on a variety of topics found in TV, radio, video, or live and computer-generated presentations
Understand basic idiomatic expressions used in daily life	Understand high-frequency idiomatic expressions in the context of the stimulus	Understand idiomatic expressions
Recognize vocabulary words related to familiar topics	Understand a variety of vocabulary and expressions related to familiar topics	Comprehend a wide variety of vocabulary and draw appropriate inferences based on interpretation of the passage
Understand a short series of simple directions	Understand a series of directions	Understand a series of steps to complete a task
Recognize differences between formal and informal language	Recognize differences between formal and informal language, plus specific expressions used for certain circumstances	Demonstrate some use of cultural context to deduce meaning of unfamiliar vocabulary

Table 7. Presentational Mode: Writing-Based Tasks

ACHIEVEMENT LEVELS		
Basic	**Proficient**	**Advanced**
Students at this level demonstrate partial mastery of the knowledge and skills needed to write simple messages in a foreign language.	*Students at this level demonstrate a solid mastery of the knowledge and skills needed to write simple messages in a foreign language.*	*Students at this level demonstrate superior performance in writing in a foreign language.*
Students functioning at the *basic* level should be able to:	Students functioning at the *proficient* level should be able to:	Students functioning at the *advanced* level should be able to:
Use lists of words as well as short, memorized phrases and sentences on familiar topics	Use simple sentences and strings of related sentences	Compose one or more cohesive paragraphs with topic sentences and supporting sentences; Vary sentence length and sentence structure (e.g., compound/complex)
Be understood by a reader very accustomed to interacting with language learners	Write about familiar topics with sufficient accuracy so that a reader with language learning experience may understand the message	Write with a fairly high degree of facility so that a native speaker will understand.
Demonstrate ability to use capitalization and punctuation conventions, although with frequent native language interference	Use capitalization and punctuation with fairly good accuracy	Use appropriate capitalization and punctuation
Express ideas using vocabulary limited to common objects and actions (many gender errors will be evident)	Select a variety of task-appropriate vocabulary and some idiomatic expressions (although may use occasional false cognates and make occasional gender errors)	Use an extensive vocabulary including a number of correctly used idiomatic expressions
Demonstrate some limited control of verb forms	Write in the present time accurately	Manipulate verbs (tense, voice) with no pattern of errors
	Use an occasional transitional word	Use connecting/ transitional words to relate sentences
	Paraphrase to avoid difficult syntactic structures or unfamiliar vocabulary	Express ideas with sufficient clarity so that little or no paraphrasing is needed
	Use some culturally appropriate writing conventions	Demonstrate control of culturally authentic expression and writing conventions
	Demonstrate some awareness of target audience perspective/ practices	Demonstrate increasing awareness of target audience perspective/ practices

Table 8. Interpersonal Mode: Conversation-Based Tasks

ACHIEVEMENT LEVELS		
Basic	**Proficient**	**Advanced**
Students at this level demonstrate partial mastery of the knowledge and skills needed to participate in simple conversations in a foreign language.	*Students at this level demonstrate a solid mastery of the knowledge and skills needed to participate in simple conversations in a foreign language.*	*Students at this level demonstrate superior performance in participating in many kinds of conversations in a foreign language.*
Students functioning at the *basic* level should be able to:	Students functioning at the *proficient* level should be able to:	Students functioning at the *advanced* level should be able to:
Understand very simple questions and comments on familiar topics; may need restatements or gestures in order to understand a message	Understand and respond to what others say to them on familiar topics, asking for clarification when necessary	Understand and respond to what others say to them on a range of topics
Express themselves in single words, short phrases, or a simple sentence	Express themselves in several consecutive sentences	Express themselves amply and comfortably, with elaboration as needed to make a point
Use words and phrases produced by the interviewer to express themselves	Make an effort, sometimes successful, to express something for which the needed vocabulary and/or structures are not known	Show resourcefulness to successfully express something they do not know how to say
	Use words, phrases, and structures (primarily present tense) that are appropriate to the topic and to the conversational situation	Show accuracy and fluency when narrating and describing in present time; show some accuracy in past and future times
	When asked to clarify something they have said, sometimes offer an alternative way of expressing the thought	Easily clarify and expand upon something they have said when asked to do so
	Ask questions to get information	Ask questions to get information and clarify something that has been not clearly understood

Exercise Format and Content Specifications

For the 2003 Spanish NAEP, each exercise will contain a specific stimulus selected from a variety of materials and a communication task requiring a student response in one of four response formats: multiple-choice, written short constructed response, written extended constructed response, or oral performance.

Stimulus Materials

In the development of all types of exercises, it is important to control the stimulus tightly. The exercise developer must formulate clearly in his/her own mind the nature of the task the student is to perform (i.e., the kind of operation the student is intended to carry out); define the materials on which the operation is to be carried out; and incorporate into the exercise a clear statement of the task to be performed. The statement must include the situation or setting the examinees are to be given and what they are to do.

Stimulus material must be varied and can include spoken, written or other non-textual types of materials. Allowable spoken materials include dialogues; interviews; directions; radio commercials, announcements, and reports; stories; and other readings. Written materials can include stories; information passages; dialogues; directions; menus; newspaper and magazine articles or advertisements; and bus, train, or plane schedules. Allowable non-textual stimuli accompanying these spoken and written materials include photos, artwork, graphics, and videos. While a variety of types of materials should be used, the complexity of the materials used should be appropriate for students at the twelfth grade level. In addition, careful consideration should be given to the amount and level of reading and listening material presented in the stimuli for the interpretive questions. Stimulus materials should conform to specific guidelines given in the following sections.

Response Formats

The response formats to be used for exercises in the Interpretive mode include multiple-choice and short answer written constructed response. For the Presentational mode, response formats include short answer and extended written response. Multiple-choice exercises will be of the conventional single correct option format with four options and will ask students to select the response that best answers the question. The short answer constructed response exercises will require the students to respond in short answers that may vary from one or two words or phrases to several sentences. Extended response exercises will require students to respond with written responses of a paragraph or more.

For the Interpersonal mode, the response format will be oral performance (a conversational interaction). The interpersonal conversation exercises will require the students to participate in social and role-play conversations with the administrator using words, phrases, sentences or even paragraph length discourse. Both the test administrator and the student will be asking and answering questions as part of the conversation. In addition, the students may be asked to fill out a written form (part of the prompt) to indicate completion of the task. This form should not be scored.

Detailed specifications for exercises to be developed in each of the communication modes to be assessed on the Spanish NAEP follow.

Assessment Tasks in the Interpretive Mode

The *interpretive* mode refers to the understanding of spoken or written language, such as listening to a broadcast or film, or reading a magazine or short story. It involves understanding, in a culturally appropriate way, the meaning of oral or written messages presented via print, audio, and visual (such as video) materials. In this mode, the originator of the message is not present. There is no "give and take" between participants, and misunderstandings cannot be clarified directly. Necessary for successful communication in this mode are the receptive language abilities of listening and reading.

For the purposes of the Spanish NAEP assessment, communicative ability in the interpretive mode will be assessed through two broad types of tasks: reading-based tasks and listening-based tasks.

Allowable Response Formats

The response formats used for the exercises in the interpretive mode include multiple-choice and short written constructed response. Multiple-choice exercises will be of the conventional single correct option format with four options and will ask students to select the response that best answers the question. Constructed response exercises will be used to provide insights into students' understanding of Spanish. These exercises will require students to express their understanding and ideas in English. The short written constructed-response exercises will require students to respond in a few words or several sentences. The consensus building committees recommend that, for both listening- and reading-based tasks, an appropriate mix of multiple-choice and short answer items be used for each level of the assessment so that the proportion of response formats is 50:50 in each of the total Listening and Reading item pools.[3]

The Interpretive Mode: Reading-Based Tasks

Purpose

In the *NAEP Reading Framework*, there are three purposes for reading. These include reading for literary experience, reading to perform a task, and reading for information (NAEP Reading Consensus Project, n.d.). For the Spanish NAEP, the main purpose for reading is to acquire information for understanding and application. This purpose views reading as a process by which the reader must understand the meaning of texts written in the Spanish language in a way that is appropriate to the culture in which the text was produced. This view of reading moves beyond simple literal comprehension of the foreign language writing to the ability to interpret the text.

[3] If there are only two levels of the assessment, the committees recommend that the distribution be 60% multiple choice and 40% short answer for the lower level (Level I), and that the percentages be reversed for the upper level (level II).

An additional element in the Foreign Language NAEP Reading requires consideration of the ability to understand and apply cultural contexts and references from the text.

On the Spanish NAEP, all reading assessment tasks should be placed within one of three contexts: daily living, school, and work. Assessment tasks should provide students with contextual information about the purpose(s) for reading, for example, how the information gained from the text will be used and should indicate the context in which the written text is found, such as an excerpt from a magazine article or a sign in a store.

Topics

All content must be appropriate to 12th grade students and relate to daily living, school, and work. Examples of allowable daily living topics include leisure activities, routines, friendship, homes, community, and significant occasions. Examples of academic topics include geography, history, health, music, art, and the environment. Topics related to work include getting jobs, understanding work guidelines, and comparing work practices in other cultures. Note that topics concerning the test-taker's personal or family life (including home and living arrangements) are considered sensitive and must be avoided.

Text Characteristics

Appropriate written/visual texts should be typified by the following characteristics. They should:

- vary in length from 25 to 600 words;
- contain some cognates (i.e., words in English and Spanish that look similar and have the same meaning, for example the English word science and the Spanish word ciencia), redundancies, and high frequency idiomatic expressions;
- contain familiar vocabulary related to daily life, school, and work;
- be non-technical in nature;
- show simple, sentence-level and/or connected discourse;
- be in a familiar script if other than typed;
- display a high level of cohesion;
- in longer texts, be organized as description, narration, or expository text, containing a mixture of time frames and hypothetical language at times;
- display a functional use of language to express or exchange information about ideas, knowledge or feelings (such as descriptions and explanations) or to influence or affect the world (such as requests, commands, greetings, compliments);
- contain Spanish from Spain and Latin America as used in printed media, but avoid slang and language that is limited to a certain local area or region;
- contain register (i.e., the style of language appropriate to the specific social setting) that is formal or informal as related to the school, work or daily life situation;
- contain language that is natural;
- have imbedded cultural information about perspectives, practices, and products at times that lend themselves to student interpretation; and
- contain judicious use of figurative language, if at all.

Sources of appropriate texts include advertisements, surveys, newspapers and magazines geared toward teenagers (such as *Hola, Tiempo, El Nuevo Herald, Cambio 16, El Tribuno, El Pais*, and

ABC), textbooks, brochures, travel books, Internet home pages and Web sites, film reviews, TV guides, biographies, *People* magazine in Spanish, poems, and literary excerpts.

Performance Dimensions

Multiple choice and short constructed response items in English should allow students to meet the evaluative criterion of demonstrating comprehension of Spanish texts in this mode. Items that allow students to demonstrate comprehension include those that ask students to do the following:

- provide an initial impression or global understanding of the text;
- extend initial impressions to develop more complete understanding;
- link information across parts of a text as well as focus on specific information; and
- use critical evaluation strategies to compare/contrast and understand the impact of cultural features (products, practices and perspectives) contained in the text.

Items that allow students to demonstrate the use of communication strategies include those that ask students to do the following:

- infer meaning of unfamiliar words and structures using multiple clues (i.e., contextual, linguistic, formatting, visuals); and
- use graphics, titles, and visuals to assist with meaning.

Items that allow students to demonstrate application of cultural knowledge include those that ask students to do the following:

- use knowledge of their own culture and the cultures of Spanish-speaking countries and communities to interpret texts; and
- identify relationships among perspectives, practices, and products of the cultures of the Hispanic world.

Differences between lower level and upper level tasks should be based on several dimensions. Lower level tasks will tend to have texts that are shorter in length. Lower level texts should not contain a high degree of abstraction (e.g., symbolic language and metaphors). In lower level texts, structures alone should not convey the meaning. Finally, the nature of the task itself may determine whether it is suitable for an upper level or lower level of the assessment. In other words, at times the same text may be used for generating upper level and lower level tasks.

Integration with the Four Goal Areas

Tasks integrated with the *cultures* goal should be based on texts containing information related to practices, products, and/or perspectives of cultures of the Hispanic world such as traditions, holidays, school life, family life, housing, and sports. Content in such texts should be culturally authentic; that is, it should contain real language used by real Spanish-speaking people, not be texts written for teaching non-native Spanish speakers.

Tasks integrated with the *connections* goal should be based on texts related to academic topics such as health, geography, history, everyday science, environment, music, art and provide opportunities to learn new information through Spanish.

Tasks integrated with the *comparisons* goal should be based on texts that present features of the language and/or cultures of Spanish-speaking countries and communities and ask students to reflect on how they compare to their own. Such texts should allow students to see similarities and differences between their own language and culture and those of Spanish speakers.

Tasks integrated with the *communities* goal should be ones in which the students read a text that allows them to see the use of language outside the classroom (e.g., texts about community-based projects and exchange programs, or cultural events in the Hispanic community) or for which the purpose for reading the text exists outside the classroom, such as for personal enrichment or enjoyment.

The Interpretive Mode: Listening-Based Tasks

Purpose

There are many purposes for listening. Although there is no framework for a NAEP in listening, at least three purposes for listening can be identified, parallel to those in the NAEP reading framework: listening for literary experience, listening to perform a task, and listening for information. For the Spanish NAEP assessment, the main purpose is listening for information, both to understand it and to know how to use it. This purpose views listening as a communicative interaction in which the listener must understand the meaning of the message within its cultural context. This view of listening moves beyond literal translation of foreign language speech.

Context for Listening

All listening assessment tasks should be placed within one of three contexts: daily living, school and work. Assessment tasks should indicate the situational context for the audio/visual passage, for example, an excerpt from a television broadcast or a message on an answering machine. Tasks should inform students about the purpose(s) for listening, for example, how the information gained from the text will be used. In order to focus on the purpose of listening, students should see the exercise questions or tasks before the listening passage begins. How often the prompt is heard should be determined by the nature of the authentic listening task. For example, in real life a message left on voice mail can be replayed, so voice mail messages on the Spanish NAEP, for example, should be heard at least two times. For weather reports, stories, television dramas, and interviews, the voices are not heard a second time; however, redundancy is a natural component of such situations. For passages of this type and for conversations, the listening passage should be heard only once, but paraphrasing and re-statement should be integrated into the passage. Note taking, as done in real life (for example, for phone messages) should be allowed for listening passages and tasks to which it is appropriate in real life.

Topics

All content must be appropriate for twelfth graders and relate to daily living, school, and work. Examples of allowable daily living topics for the listening tasks include family, leisure activities such as hobbies and entertainment, home and community, weather, social routines such as meals and birthdays, and holiday celebrations. Examples of school topics include school life and activities, and academic themes such as geography, history, health, music, art, science, and

language study. Examples of work topics include situations that 12th graders would encounter in the work place (office/clerical, service/repair, and ads/interviews) or in the volunteer environment (hospital, civic or service organization, club).

Passage Characteristics

Appropriate audio/visual passages should:

- be of a length that will not overtax memory (5 – 90 seconds)—passages may be shorter for lower level tasks and somewhat longer for upper level tasks;
- be recorded at moderate conversational speed by native speakers whose speech is clearly articulated;
- consist of a series of short to medium-length sentences;
- contain minimal use of regional accents and vocabulary;
- be based on authentic language materials from Spanish-speaking cultures;
- be appropriate to the societal and cultural norms of Spanish-speaking countries/communities;
- express or exchange information about ideas, knowledge or feelings, such as descriptions, explanations, requests, commands, greetings, and compliments;
- contain formal and informal styles of language that are appropriate to the specific social setting of the passage;
- contain authentic cultural references; and
- use everyday language.

Sources of appropriate listening texts need to be authentic or based on an authentic purpose. From television, allowable sources include documentaries, talk shows, news programs, and advertisements. From radio, allowable sources include news, weather reports, and talk shows or interviews. Conversational passages for use as listening texts should be generated by giving a topic and situation to two or more native speakers and asking them to have a conversation on that topic. The speakers then create the listening text using their natural conversation techniques and following guidelines provided by the item writer for level and length of conversation, rather than reading from a script.

It is suggested that 50% of the Spanish listening passages be presented through audio and 50% be provided through video. These passages may best be presented through a computer-based format. In a lower-level task, audio may be used for announcements, phone conversations or messages, or radio weather reports. Video might be used for the telling of a story or other passages where visual support would help set the context without giving away direct comprehension answers. For a higher-level task, audio might be used for radio clips, native speakers' conversations, and phone conversations. Video can be relied on for material that is more authentic, intended for native speakers, but still accessible to higher ability foreign language students.

Performance Dimensions

Students will demonstrate listening comprehension through multiple-choice and short constructed response items in English. Items assessing comprehension should ask students to

- identify main ideas;
- identify specific information;
- understand basic idiomatic expressions and vocabulary used in daily life; and
- draw appropriate inferences (at advanced levels).

Items that allow students to demonstrate use of communication strategies to aid comprehension should ask students to

- use cognates (i.e., related words in English and Spanish that sound similar and have the same meaning, such as English word science and the Spanish word ciencia);
- use context clues (and rely on visuals to set the overarching context at the basic level);
- use basic structures within familiar contexts; and
- use tone, pitch and emotion.

Items that allow students to demonstrate application of cultural knowledge should ask students to

- use knowledge of their own culture and of those of Spanish-speaking countries and communities to interpret passages;
- identify relationships among perspectives, practices and products of cultures of Spanish-speaking countries and communities;
- demonstrate understanding of both formal and informal language usage; and
- demonstrate understanding of use of emotion and tone in speech patterns.

Integration with the Four Goal Areas

Tasks integrated with the *cultures* goal should be based on listening texts containing information related to cultural practices, products and/or perspectives whose content is authentic (i.e., real language used by native speakers, not contrived for pedagogical purposes of non-native speakers). Authentic texts may show use of both formal and informal language in appropriate contexts.

Tasks integrated with the *connections* goal should be based on listening texts related to academic topics such as health, geography, environment, music, art, and should provide opportunities to learn new information in Spanish.

Tasks integrated with the *comparisons* goal should be based on listening texts that present features of the Spanish language and/or the culture of Spanish-speaking countries and communities and should ask students to reflect on how they compare to their own. Such texts should allow students to see similarities and differences between their own language and culture and those of Spanish-speaking countries and communities. They may demonstrate cultural uses of both formal and informal language.

Tasks integrated with the *communities* goal are ones in which the students listen to a text that allows them to hear the use of language outside the classroom (e.g., texts about community-based projects and exchange programs) or for which the purpose for listening to the text exists outside the classroom, such as listening for personal enrichment or enjoyment. The setting for such a task may be from a Spanish-speaking community within the U.S.

Assessment Tasks in the Presentational Mode

The *presentational* mode involves spoken or written communication, such as giving a speech or writing a story. It involves producing spoken or written messages for an audience with whom there is no immediate personal contact. In this mode, there is no possibility to clarify one's intended meaning if the audience does not understand it. Such messages need to reflect awareness of cultural differences and an attempt to bridge those differences in order to enable appropriate interpretation by persons with a cultural background from a Spanish-speaking country or community. Necessary to achieving successful communication in this mode are the productive language abilities of speaking and writing and the ability to use visual images to add in making oneself understood.

For the purposes of the Spanish NAEP assessment, communicative ability in the presentational mode will be assessed only through tasks that require the student to communicate in writing.

The Presentational Mode: Writing-Based Tasks

Allowable Response Formats

The response formats used for the exercises in the presentational mode include written constructed response, both short answer and extended response. Constructed response exercises will be used to assess student ability to make a written presentation in Spanish based on a stimulus presented in English. The written constructed-response exercises will require students to respond in words, phrases, or sentences to extended responses of a paragraph or more. Extended-response exercises in the presentational mode will ask students to write paragraph-length or longer responses that demonstrate their ability to communicate in Spanish. Students will be told what will be scored (i.e., what will is expected) through directions that will help them think, plan and write for each task. These may be in the form of questions that help them focus on the task.

There should be a mix of short- and extended- constructed response exercises in the exercise pool. To best utilize students' time and gain the most information, it is suggested that there might be more short constructed response exercises at the lower level of the assessment. If an extended response is required, the student must have 25 minutes to respond. If a task requiring a short constructed response also is included, the time will have to be extended.

Purpose

In the NAEP Writing Framework, there are three purposes for writing. These purposes are narrative writing (producing stories or personal essays), persuasive writing (writing to influence others to take some action or bring about change), and informative writing (writing to inform).

For the Spanish NAEP assessment, the main purpose for writing is to provide a reader with information by sharing knowledge, descriptions, instructions, and ideas. For the Spanish NAEP, such communication should be based on content that is appropriate for 12th graders and familiar to a broad range of students from varied backgrounds. Nevertheless, writing in another language to persons from a different cultural background requires the author to be sensitive to that cultural background so that his or her writing may be properly understood.

Context for Writing

On the Spanish NAEP, all writing assessment tasks should be placed within one of three contexts: daily living, school, and work. Assessment tasks should inform students about the purpose(s) for writing the message, (for example, why the information to be communicated is necessary). Assessment tasks also should provide the situational context in which the message is to be written, including a description of the audience that will be receiving the message.

Topics

Content for task completion should be appropriate for 12th graders and familiar to a broad range of students from varied backgrounds. Topics should relate to daily living, school, and work. Examples of allowable personal themes include description of friends, important people, physical characteristics (city, countryside, school), free time activities and interests, weather, foods, daily routine. Examples of cultural themes include holidays, celebrations, foods, traditions, daily schedules, shopping, and customs, including writing conventions. Examples of academic themes include course-taking patterns in high school/junior high, school schedules and subjects, extracurricular activities, relationships to teachers, favorite subjects. Note that topics concerning the test-taker's personal or family life (including home and living arrangements) are considered sensitive and must be avoided.

Performance Dimensions

Presentational writing tasks can range from preparing lists to one-sentence instructional messages to paragraphs reporting on experiences or providing descriptions. Messages should be written by the student to native Spanish-speakers from or in a country other than the United States.

Expectations for student performance shall be made clear to the students. As in the English writing NAEP, directions must be given to students to help them with planning and pre-writing activities. Such directions may include a checklist of questions to help them focus their response to the writing task.

In terms of demonstrating comprehensibility, written messages should be evaluated on language usage, such as appropriateness of the language to the task, total number of words or sentences, length and/or complexity of sentences, vocabulary (appropriate and varied), and use of connecting words. The use of communication strategies may be demonstrated by the use of cognates (i.e., words in English and Spanish that look and sound similar and have the same meaning, such as the English word *science* and the Spanish word *ciencia*), paraphrasing and self-correction. Application of cultural knowledge may be demonstrated through the use of appropriate register (i.e., the style of language appropriate to the specific social setting) and

evidence of understanding of differences between American culture and the cultures of Spanish-speaking countries and communities.

Integration with the Four Goal Areas

Tasks integrated with the *cultures* goal should provide opportunities for students to explicitly discuss cultural similarities/differences.

Tasks integrated with the *connections* goal should relate to impersonal or academic topics such as the food pyramid, health, education, geography, habitat, weather, and distances.

Tasks integrated with the *comparisons* goal should allow students to compare and contrast, for example, school, climate, daily schedule, and holiday customs with those in Spanish-speaking countries.

Tasks integrated with the *communities* goal should allow students to use writing outside of the school setting and as part of becoming life-long learners by using the language for personal enjoyment and enrichment.

Assessment Tasks in the Interpersonal Mode

The *interpersonal* mode involves two-way, interactive communication, such as conversing face-to-face or exchanging e-mail messages. It is characterized by direct communication between individuals who are in personal contact, thus allowing the participants to clarify meanings when misunderstandings occur. Participants in a communicative interaction can use both linguistic and non-linguistic feedback to ascertain the extent to which their message is being successfully communicated, and can make adjustments and clarifications accordingly. Successful communication in the interpersonal mode includes the productive language abilities of speaking and writing; the receptive abilities of listening and reading; and in face-to-face interactions, the ability to use and interpret body language.

For the purposes of the Spanish NAEP assessment, the interpersonal mode will be assessed through five conversation-based tasks. The assessment of interpersonal conversation is based on the Oral Proficiency Interview (OPI), a "a standardized procedure for the global assessment of functional speaking ability" (Swender, 1999, p. 1). The OPI assessment procedure has been in use since the 1950s. Originally developed by the Foreign Service Institute to assess the speaking proficiency of persons serving in the United States Diplomatic Corps, the procedure became widely used across government agencies. In the late 1970s and early 1980s, oral proficiency testing was introduced to the academic world through the work of the Educational Testing Service and the American Council on the Teaching of Foreign Languages. Through ACTFL's efforts, thousands of educators have been trained in the procedure. The ACTFL OPI is recognized world-wide by hundreds of institutions of higher learning, government agencies, State Teacher Licensing Boards and Fortune 100 corporations as a proven method of assessing oral language proficiency. OPI ratings are used for a variety of purposes, among them the awarding of college credit, the credentialing of individuals for work and study opportunities, the determination of hiring and promotion policies in the workplace, and the evaluation of the effectiveness of instructional programs in the United States and abroad. Language Testing

International (the ACTFL Testing Office) conducts thousands of interviews each year. In the experience of government, academia, and business, the OPI has shown itself to be a valid and reliable assessment technique.

The FL NAEP will retain the interactive nature of the OPI as a means of eliciting a speech sample that enables students to speak to the best of their abilities. However, to ensure the standardization necessary for the purposes of the FL NAEP, the OPI procedure will be modified in several ways. First, while in the OPI the topics of the interview can vary according to the interests and experiences of the examinee, for the NAEP the topics have been standardized to cover three prescribed contexts (daily life, school, work). Second, initial and follow-up questions in the NAEP are drawn from a prepared set of standardized questions that will be common across interviews and the procedure for selecting follow-up questions will also be standardized. In addition, there will be a limited set of standardized role-play tasks. All questions and tasks will be field tested to ensure their appropriateness, the reliability with which they can be scored, and their validity as assessments of interpersonal communication. Third, the scoring of the tasks for the FL NAEP uses rubric similar to those used in other NAEP assessments, rather than the global proficiency level descriptors used to rate the OPI. The use of the rubrics will further enhance the reliability of the scoring of the tasks. In addition, as mentioned elsewhere, the thorough training of interviewers and raters involved in the NAEP will further ensure its standardization. The above modifications will help to standardize the procedure across students, while ensuring that the tasks are valid and reliable for the purposes of the NAEP in terms of administration and scoring.

The Interpersonal Mode: Conversation-Based Tasks

Allowable Response Format

For exercises in the interpersonal mode, the response format is a conversational interaction; i.e., an oral performance. Each exercise requires a student to participate in a multiple task-based conversation with the administrator using words, phrases, sentences, or even paragraph length discourse. Both the test administrator and the student will ask and answer questions as part of the conversations.

Purpose

There are many purposes for participating in conversations, such as exchanging information, establishing and maintaining social relationships, and accomplishing social transactions. In the Spanish NAEP, examinees will participate in a conversation for two purposes: 1) to establish rapport with an interviewer and 2) to exchange information with the interviewer to accomplish specific tasks. One participant will be the student, the other a fluent speaker of Spanish, trained to conduct oral interviews in Spanish. The conversation will consist of two phases. In phase 1, the social conversation phase, a relationship between the participants will be established through a conversation on familiar topics. This phase will consist of three tasks. In phase 2, the student and the interviewer will work together to accomplish two role-play tasks. The content for both phases must be appropriate for twelfth graders and familiar to a broad range of students from varied backgrounds.

Contexts for the Social Conversation and Role-play Tasks

On the Spanish NAEP, daily living, school, and work will provide context for three social conversation tasks and for two role-play tasks. In the social conversation phase, the instructions should indicate how the interviewer and the student are to establish rapport. For the role-play tasks, the instructions should provide all necessary details for the situational context in which the conversation takes place. In phase 1, the student and interviewer are themselves, as they are establishing rapport in this phase. In phase 2, the student and interviewer assume roles to carry out the role-play tasks. A more detailed description of the two phases is provided under Guidelines for Exercise Construction, below.

Topics

As mentioned above, content for task completion should be appropriate for 12th graders and familiar to a broad range of students from varied backgrounds. Topics should be related to daily living, school, and work, and be of a personal, cultural or academic nature. Technical topics should be avoided. Themes related to personal topics include daily activities/routines, leisure activities, interests and hobbies, community/town description, school as a social environment, shopping, friends, and work /jobs. Topics to be avoided are home environment, family, and dating. Themes related to cultural topics include celebrations, holidays, music and the arts in general, community and social activities, and family issues (e.g., roles and typical adolescent behavior). Themes related to academic topics include school subjects, career plans, academic interests, school projects/activities, description of school building, and the education system/structure. Note that topics concerning the test-taker's personal or family life (including home and living arrangements) are considered sensitive and must be avoided.

Performance Dimensions

For the phase 1 tasks, the interviewer will draw from an established inventory of questions or prompts to encourage the student to talk about daily activities, school, and work to the best of the student's ability. In the phase 2 (role-play) tasks, the interviewer and the student will assume roles and will exchange information to produce a set of suggestions, arrive at a plan of action, make a decision, and the like. The student may be asked to fill out a written form (part of the prompt) to indicate completion of the task, but the form should *not* be scored. Although, the task instructions may imply that the student and the interviewer are to participate equally in the interaction; however, the interviewer should encourage the student to do most of the talking. Both phase 1 and phase 2 tasks will be scored with holistic rubrics containing 4-6 score points. The five task scores will be aggregated for a single interpersonal communication score.

In both the social conversation and the role-play tasks, the interviewer should:

- speak in Spanish only;
- participate in the interaction to elicit maximum student speech and when possible place the burden of maintaining the interaction on the student;
- promote interaction and negotiation between the interviewer and the student;
- make gestures when appropriate to the interaction;
- use vocabulary that is high-frequency and appropriate to the task;
- use vocabulary and pronunciation that would be universally understood by Spanish speakers and not specific to any particular region;

- adapt complexity of discourse within the task to performance level of student (i.e. tense, syntax, vocabulary);
- use formal or informal speech as appropriate to the role-play task; and
- use non-verbal behavior that is non-threatening and shows interest in what the student is saying.

Tasks in both phases should allow students to produce speech that is characterized by:

Comprehension

- participation in the conversation in a way that indicates that the student understands the interviewer's utterances, and is able to request repetition or clarification when the interviewer's meaning is not fully understood;
- appropriate response to requests for clarification, expansion, or elaboration by the interviewer; that is, response to these requests that indicates that the requests have been understood; and
- demonstration of appropriate turn-taking behavior (i.e., knowing when the interviewer has finished his or her turn).

Comprehensibility

- use of Spanish language only; or effort expended in trying to maintain discourse in Spanish;
- production of utterances that are of appropriate length in the context of the interaction;
- control of high-frequency vocabulary;
- ability to use vocabulary that is appropriate to the task;
- control of the syntactic structures needed for face-to-face conversation on familiar topics (e.g., present tense, question formation, noun-adjective agreement); some use of more advanced structures;
- use of Spanish sound system in a way that a native Spanish-speaking interviewer who has had prior contact with learners of Spanish can understand easily;
- moderate control of functional language (i.e., speech acts) [see also Cultural Knowledge section below]; and
- demonstration of knowledge of when to speak and when to remain silent, when to pause and how to pause.

Communication Strategies

- appropriate use of connectors and other function words that link utterances cohesively;
- appropriate conversational turn-taking behavior;
- ability to deal effectively with breakdowns in communication;
- use paraphrase to express himself or herself when lacking a specific word; and
- use clarification and verification to check on appropriateness or comprehensibility of his or her utterances.

Cultural Knowledge

- selection of speech acts that are appropriate to the context (i.e., knowing when to make a request, offer an apology, make a compliment, open and close an interaction);
- demonstration of ability to adjust the level of formality that is appropriate to the task (e.g., appropriate use of tú and usted);
- demonstration of cultural knowledge by making explicit cultural references according to the task (e.g., asking about appropriate clothing for a trip to the capital city Caracas); and
- ability to ask appropriate questions to get cultural information.

Integration with the Four Goal Areas

Tasks integrated with the *cultures* goal could include a role-play that involves doing a joint project on a cultural topic (e.g., Hispanic cultural practice, country, or person). Tasks integrated with the cultures goal should provide opportunities for students to explicitly discuss cultural similarities/differences with the native Spanish speaker.

Tasks integrated with the *connections* goal should focus on a country or topics such as science, arts, or a famous person.

Tasks integrated with the *comparisons* goal standard should involve those in which participants talk about own culture versus that of others.

Tasks integrated with the *communities* goal should allow students to participate in conversations outside of the school setting and as part of becoming life-long learners by using the language for personal enjoyment and enrichment.

Guidelines for Exercise Construction

Multiple-Choice Exercises

Good multiple-choice items will be constructed to test students' abilities to understand spoken and written stimuli in Spanish. Multiple-choice items have the following requirements for item options:

- there must be only one clearly identifiable correct option for each question; if plausible arguments can be made for more than one option, the item is unacceptable;
- the correct option must be a concise answer that will satisfy any qualified judge as being an adequate short answer to the question. The response must not answer more than the stem question asks;
- distractors (incorrect responses) should appeal to some kind of misinterpretation, predisposition, unsound reasoning, or casual reading;
- distractors should be plausible and homogeneous. Options that are obviously wrong (e.g., because of some mismatch with the stem) or silly effectively reduce the number of possible correct answers and, thus, reduce the validity of the item. Sources of good distractors include common misinterpretations and errors in reasoning, statements that are

true but that are not correct answers to the questions posed in the item stems, statements that are either too broad or too narrow to be correct, and carefully worded incorrect statements that may sound plausible to the uncritical reader;

- distractors must be written with as much care and precision as the correct option so that all alternatives are equally attractive to an examinee who guesses. Each option should be a separate and distinctly different response to the stem. Responses should not overlap or include other responses. "All of the above" should not be used as an option; and
- both the stem and the options should be as brief and straightforward as possible. All options should be parallel in point of view and grammatical structure and similar in length. The stem should include any words that otherwise would have to be repeated in each option.

Constructed Response Exercises

Constructed response exercises, requiring short and extended written responses, will be used to provide insights into students' understanding of Spanish interpretive listening and reading stimuli, as well as to assess their ability to make a written presentation in Spanish based on a stimulus presented in English. These exercises will require students to listen or read carefully and organize and express their thoughts—in English for the interpretive reading/listening questions and in Spanish for the presentational writing questions.

Responses to short-answer questions may be single words or phrases or they may be lists of words or phrases. They also may be a few sentences or a brief paragraph. Extended-response exercises will ask students to write paragraph-length or longer responses that demonstrate their ability to communicate in Spanish.

Care must be taken on all exercises to identify the response elements that should be present to constitute a satisfactory answer; that is, the stimulus should be defined (scaffolded) to indicate to the student how the exercise will be scored. Directions to the student must be complete, containing enough information to ensure that all students understand the task in the same way. This information will also explain to the student how the exercise will be evaluated, aid in the scoring process, and ensure correct interpretation of the answer.

Interpersonal Communication Exercises

Students will *not* be routed into different levels of the assessment for assessing interpersonal communication. There will be only one level of the interpersonal assessment. The methodology described below outlines the standardized administrative procedures for the interpersonal communication exercises. These procedures coupled with the focused holistic scoring rubrics will distinguish among multiple levels of proficiency on interpersonal communication tasks.

For the social conversation tasks (phase 1), the interviewer will draw from an established inventory of questions or prompts to encourage the student to talk about daily activities, school, and work to the best of the student's ability. This phase of the assessment will consist of three independent conversations (tasks). The interviewer will begin with an initial question and will help the conversation continue through the use of follow-up questions that will be adapted within a standardized, pre-set, limited inventory. There will be three fresh starts on three different topics

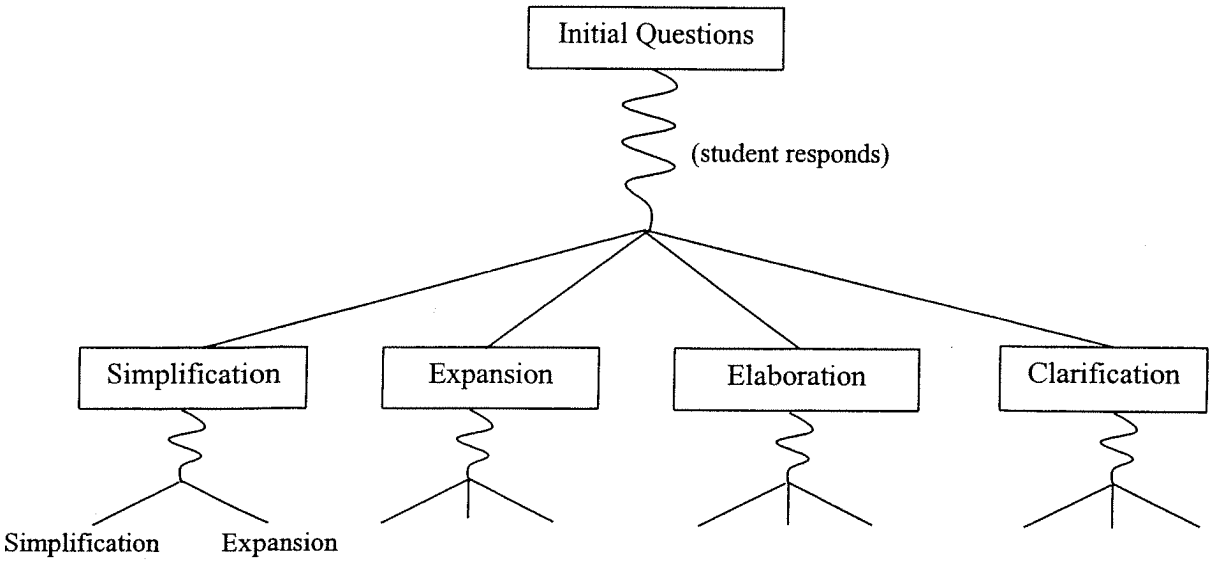
within the contexts described above. Each conversation will be evaluated with a holistic rubric. The follow-up questions are to gather more information for each rating and should be consistent with the scoring rubric. Each phase 1 conversation (task) should take only 2-3 minutes, for a maximum total time of 8 minutes for phase 1. If the responses to the three tasks in phase 1 are characterized by the "unsatisfactory" rubric descriptors (see sample rubric in Appendix A), the interviewer will terminate the conversation.

In the two role-play tasks (phase 2), the interviewer and the student will assume roles and will exchange information to produce a set of suggestions, arrive at a plan of action, make a decision, and the like. As needed, the student will be asked to fill out a written form (part of the prompt) to indicate completion of the task, but the form will *not* be scored. As mentioned above, while the task instructions may imply that the student and the interviewer are to participate equally in the interaction, the interviewer should encourage the student to do most of the talking. The interviewer must be careful to encourage the examinee to ask questions to complete the task.

Each phase 2 conversation (task) should take about 6-8 minutes. As in phase 1, the interviewer will begin with an initial question and will encourage the conversation through the use of a standardized, limited inventory of follow-up questions. Each of the phase 2 tasks will be evaluated with a holistic rubric.

For both phases (all five tasks), the follow-up questions should be consistent with the scoring rubrics and address the components of the rubrics. Specific questions will be up to the administrator within a standardized protocol and will serve the following purposes: simplification, expansion, elaboration, and clarification. When simplification is necessary, the interviewer should ask one or more simpler, more directed questions. When expansion is desired, the interviewer should ask questions to elicit more detail, a greater quantity of information, or a longer response. To obtain elaboration of a student response, the interviewer should ask questions to get the student to introduce a new linguistic function at a higher level, such as compare/contrast or narrate in past time. The interviewer should use clarification questions to elicit paraphrase. For examples of follow-up questions, see the sample interpersonal task in Appendix A. A set of purposeful questions (simplification, expansion, elaboration, or clarification) is referred to as a stage within the conversation. Figure 5 illustrates the possible follow-up stages of each conversation.

Figure 5. Stages of Follow-Up Questions



Because of time constraints, there must be a limited number of conversational stages within each of the five tasks. For phase 1, an administrator should limit each stage to a maximum of one minute or two questions, whichever comes first. In phase 2, each stage can be longer—three minutes or 3-6 questions. The maximum total time for the interpersonal conversation is 25 minutes.

The Technical Advisory Panel recommends that the testing contractor standardize the follow-up protocols through an empirical process. The contractor should identify several possible stages (types of follow-up questions) and, through field testing, identify what types of follow-up questions best elicit more ratable performances. Descriptions of what students should be able to do in response to each type of follow-up questions, for example, the ability to expand, elaborate or clarify, should be included in the scoring rubrics.

Specifications for Scoring Guides

Refinement of the scoring criteria will be an integral and ongoing part of the assessment development process. Preliminary guidelines for drafting scoring guides (rubrics) follow. (See Scoring and Reporting Specifications for more detail about evaluation of open-ended exercises and the immediately preceding descriptions of exercise types for specific performance dimensions on which the scoring rubrics should be based.).

As mentioned previously, scoring guides should be created for the constructed-response exercises (short and extended response) and for the interpersonal conversation exercises as the exercises are being developed. After the exercises have been field tested, the scoring guides should be refined based on actual student responses. When designing scoring guides, there should be a tight match between the demands of the test question and the criteria contained in the scoring guide. That is, the exercise must be designed so that the components to be scored are quite evident. The instructions to students should communicate what is being looked for in the

students' responses. An exercise should pose a specific task for students, provide them with a focus, help them understand the criteria to be used in scoring the exercise, and indicate the scope of the expected response.

Rating scales points within scoring guides should be assigned to reflect the quality of the responses, and each point on the scoring guide should contain specific criteria for the desired response at each level. The rating scales should contain only as many score categories as can be well defined, and the categories should be sufficiently differentiated to allow raters to use the various points on the scale consistently. The description provided for each category on a scale should be clear and contain specific criteria for the desired responses. The qualities or characteristics of a response associated with each point on the scale should be identified as precisely and completely as possible.

For evaluating responses to listening and reading exercises in the interpretive area and responses to writing exercises in the presentational area, scoring guides in a textual format will be suitable, but for exercises in the interpersonal area, other formats may be needed for evaluating student responses, i.e., video clips that make clear the distinctions between rating categories for these responses. Raters then would be trained to evaluate student work in relation to these nontextual exemplars.

The consensus building committees recommend that a preliminary scoring guide be drafted for each exercise *a priori*. Descriptions of hypothetical best possible responses to the exercise would define the top scale point. Somewhat less-accomplished responses that might be expected would then define the next scale point and so on.

Once a preliminary scoring guide has been drafted, examples of student responses to the exercise gathered from field testing should be used to determine whether the guide can be applied satisfactorily. Those responses that seem to fit the various category descriptions should be identified, as well as responses that do not seem to fit any of the category descriptions. If the guide does not accommodate some responses that are judged to be plausible, then the guide should be modified so that it can handle a greater variety of reasonable responses. Additionally, the exercise may need to be revised if the review of student responses to the exercise suggests that students do not understand what is expected of them.

The scoring guides should be written in lay language and should be included in the reports summarizing student performance on individual exercises and groups of exercises. In addition to communicating the scoring criteria clearly to the raters, it will be important for the scoring guides to be written in language accessible to a wide audience of readers so that the assessment results can be readily understood.

Rater Training Specification

Selection of Raters

Raters should have background and training in Spanish language education. They should have experience with Spanish language learners and an understanding of how students learn Spanish and of what students at the grade 12 level who have had various levels of Spanish instruction know and are able to do. It is desirable that they have some experience working with heritage Spanish speaking student populations.

Training of Raters

Training is extremely important for raters of the constructed response and interpersonal exercises. The testing contractor will need to conduct inter-rater reliability checks immediately after training and again after the field test.

The training program for raters of written constructed-response and interpersonal exercises should include an orientation to the NAEP Spanish Language assessment process, a review of the various constructed-response or interpersonal exercises that they will be scoring, and a discussion of each rating scale they will use. When raters are trained to apply a rating scale, they should examine samples of student work chosen to demonstrate various levels of performance (anchor papers for written response or anchor videotapes for interpersonal exercises) so they can gain an understanding of the range of performance they are likely to encounter. They should have the opportunity to discuss the samples to help them arrive at a common understanding of the meaning of each point on a rating scale. Raters should practice applying the rating scales, and their ratings should be checked to make certain that they can apply the scoring criteria reliably and validly.

Training in cultural sensitivity and instruction not to score performance based on personal bias must be explicit. While these topics are routinely addressed in rater training sessions, given the multicultural nature of this assessment, it is even more important that they be emphasized.

Prior to scoring the presentational writing or interpersonal samples, each rater will be required to qualify by scoring at least two sets of papers or videotapes that have been previously scored by expert raters. The scores assigned by each rater to the qualifying sets will be compared to those assigned by the expert raters.

For both types of exercises, rater performance will be monitored throughout the scoring process, both statistically and by random second reading by an expert rater. Raters will be provided with statistics showing how their scores compare to raters with whom they have been paired, allowing them and their supervisors to check for any tendency toward inaccurate or inconsistent scoring. Rater reliability will be determined from first- and second-rater correlations and agreement statistics, not from adjudicated scores.

Procedures will be established to monitor the amount of rater drift that occurs during the scoring process. This is typically done through the use of validity papers that are randomly assigned to raters throughout the scoring process. The relationship between rater performance and the

"master" score assigned to the validity papers will be carefully monitored. A similar process will be established with videotapes to monitor drift that might occur during the scoring of the interpersonal exercises.

To facilitate the achievement-levels setting process, all student responses used for the achievement-levels setting process will be double scored. The performance of raters on validity papers and videotapes will be checked, and any necessary recalibration of raters will be done prior to the scoring of responses to be used in the achievement-levels setting process.

CHAPTER 3. LANGUAGE SURVEY/ BACKGROUND QUESTIONNAIRE

Background to the LS/BQ

Background questionnaires are a key part of the NAEP program. Generally NAEP has used three questionnaires: the student questionnaire, the teacher questionnaire, and the school questionnaire to gather information on instructional practices and school and classroom policies. This information has provided an important context for the reporting of achievement-level results. It also provides a description of school and teacher characteristics known to be associated with achievement, so the results may influence American education for better achievement in the future.

For the foreign language NAEP, the consensus building committees thought that much useful information about American education could be gleaned from a few well-chosen, well-measured sets of variables related to important policy issues surrounding the teaching of foreign languages. For this reason, the committees are recommending development of a language survey/background questionnaire (LS/BQ) that will be administered to a larger sample of students than those who would participate in the assessment and will include heritage language speakers of Spanish and other languages. The purpose is to gather information about students' background of experiences with languages other than their primary language, including the role that Spanish or other heritage languages play in the home environment, travel outside the country, other cultures, and so on. Students will be asked background questions about their experiences, attitudes, and knowledge about language and culture. Reasons for developing the LS/BQ include the following:

- optimize the opportunity to obtain data on a national level about a variety of factors that would contribute to a national description of second language acquisition;
- gain information about instructional experiences with second languages other than Spanish;
- assess student attitudes about foreign language instruction and experiences with second language acquisition;
- collect information about use of technology in second language instruction;
- determine perceptions of requirements for foreign language study; and
- assess students' future plans for acquisition of second language(s).

Administration of the LS/BQ

To accomplish a two-stage sampling process, the LS/BQ is to be administered to a nationally representative sample of 12^{th} grade students in each of the NAEP-sampled schools. (Schools will be sampled such that a representative sample of students in each reporting group described in Table 1 will be included in the Spanish NAEP.) Each student will take the LS/BQ at a computer (either on-line or with a disk) sufficiently ahead of the administration of the Spanish NAEP to ensure that the data collected are analyzed in time to inform the second stage of the sampling.

One purpose of the LS/BQ is to identify those students who will go on to take the Spanish assessment. The data also will include sources of information to provide accurate placement of students into one of the two or more difficulty levels of the main NAEP assessment. Sources of information may include one or more of the following: a self report of proficiency using items in which students indicate what they "can do" in Spanish, background questions (e.g., number of years of study, courses taken), school transcript information, teacher ratings of proficiency, and a short questionnaire on Spanish vocabulary knowledge. It is initially proposed that the routing procedures include a short set of questions on Spanish vocabulary as such questionnaires have been identified in research as effective proxies for estimating a general level of proficiency in a foreign language. Because using the language survey/background questionnaire in a two-stage sampling process is new to NAEP, the Technical Advisory Panel recommends that the LS/BQ be piloted on a small scale prior to the national field test.

Student Variables

Background data that are collected on the student questionnaire of the FL NAEP include those related to the following variable topics:

- demographics;
- attitudes toward language study;
- academic study and experiences with foreign language learning; and
- beyond school experiences, including the use of a foreign language in the home

Students who take the LS/BQ will represent a wide range of foreign language experiences. Some will have taken classes much earlier than 12^{th} grade. Others will have had experiences with languages other than Spanish both inside and outside of school. Because of this range of experience, it is difficult to gather the specific information on classroom experiences to present an accurate picture of foreign language teaching in the United States. The LS/BQ should be designed carefully for analyses that will allow such a picture to emerge. In particular, it is desirable to learn how much classroom instruction is based on components of the FL NAEP framework; i.e., how instruction is distributed across the modes of the framework and in what ways all five "C's"—communication, cultures, connections, comparisons, and communities—are integrated in instruction. It is also of great interest to know how much emphasis is placed on the development of communicative proficiency.

An additional facet of the LS/BQ is the student self report of proficiency in languages other than English. For Spanish students, this information may be used as part of the routing procedure. Data from the self reports of proficiency may be compared to actual achievement on the Spanish NAEP to make predictions about the achievement of other students who completed the self-report for languages other than Spanish and were not administered a FL NAEP in the foreign language they have learned.

Teacher and Instructional Variables

The purpose of the teacher questionnaire for the FL NAEP is to gather data on instructional methods and other factors related to the teacher. Information on instructional methods can provide insight, for example, into the move toward communicative language teaching or the use of national or local foreign language standards to frame instruction. Other instructional practices that are of interest include the following:

- amount of instructional time focused on each of the three communicative modes in the framework;
- description of Spanish language programs designed for heritage speakers;
- teacher's classroom use of language being taught ;
- students' required use of language being taught;
- technology, including distance learning;
- classroom assessment procedures; and
- use of curricular materials and resources

In addition to information about their teaching practices, it is recommended that the teacher questionnaires elicit information about teacher demographics (gender, race/ethnicity), preparation, (education, subject-area knowledge, language proficiency, credentials, foreign study), support for their teaching (administrative and community resources), professional development (membership in foreign language professional organizations, continuing education), and personal experience.

School and Community Variables

The FL NAEP consensus building committees recommend giving this part of the background questionnaire to the foreign language department chair at each school where students participate in the FL NAEP, rather than the school principal. For the school questionnaires, school organizational variables such as class size, availability and extent of bilingual education, whether decision-making is centralized or decentralized, and whether a district has adopted challenging content standards (despite a lack of a grammar focus) in foreign language can be explored in non-core modules. Current trends in foreign language teaching might lead one to inquire about the use of block scheduling or other service delivery approaches (e.g., intensive, immersion, or multilevel, in foreign language teaching).

School variables that are of particular interest for the FL NAEP include the following:

- school demographics (size, ethnic makeup, socio-economic status [SES] distribution, location);
- organization (public/private, length of school year, scheduling, average class or section size, governance);
- resources (financial, technology, personnel, facilities, social services);
- curriculum (standards, frameworks, course offerings);
- school culture (reform initiatives, leadership, accountability, professional development, teacher collaboration, parent involvement, orientation to foreign language instruction);

- school climate (mission, student mobility, attendance, community support); and
- school community factors (urbanicity/rurality, socio-economic status [SES] distribution, community resources).

For a more detailed description of the Language/Survey/Background Questionnaire, please see the separate document entitled *The Foreign Language NAEP Background Variables Document.*

REFERENCES

Alderson, J. C., & Urquhart, A. H. (1985). This test is unfair: I'm not an economist. In P. Haupman, R. LeBlanc, and M. B. Wesche, (Eds.) Second language performance testing/L'Evaluation de la 'performance' en langue seconde (pp. 25-44). Ottawa: University of Ottawa Press.

American Council on the Teaching of Foreign Languages. (n.d.). ACTFL Performance Guidelines for K-12 Learners. Yonkers, NY: American Council on the Teaching of Foreign Languages.

Anderson, R. C., & Freebody, P. (1981). Vocabulary and knowledge. In J. T. Guthrie (Ed.), Comprehension and teaching: Research reviews (pp. 77-117). Newark, DE: I.R.A.

Anderson, R. C., & Freebody, P. (1983). Reading comprehension and the assessment and acquisition of word knowledge. In B. Huston (Ed.), Advances in reading/language research (pp. 231-256). Greenwich, CT: JAI Press.

Beck, I. L., MeKeown, M. G., & Omanson, R. C. (1987). The effects and uses of diverse vocabulary instructional techniques. In M. G. McKeown & M. E. Curtis (Eds.), The nature of vocabulary acquisition (pp. 147-164). Hillsdale, NJ: Lawrence Erlbaum Associates.

Carrell, P. L. (1988). Some causes of text-boundedness and schema interference in ESL reading. In P. Carrell, J. Devine, & D. Eskey (Eds.), Interactive approaches to second language reading (pp. 101-113). Cambridge: Cambridge University Press.

Carroll, J. B. (1971). Development of native language skills beyond the early years. In C. Reed, (Ed.), The learning of language (pp. 97-156). New York: Appleton-Century-Crofts.

Davis, F. B. (1968). Research in comprehension in reading. Reading Research Quarterly, 3, 499-545.

Haynes, M., & Baker, I. (1993). American and Chinese readers learning from lexical familiarization in English texts. In T. Huckin, M. Haynes, & J. Coady (Eds.), Second language reading and vocabulary acquisition (pp. 130-152). Norwood, NJ: Ablex.

Koda, K. (1989). The effects of transferred vocabulary knowledge on the development of L2 reading proficiency. Foreign Language Annals, 22, 529-540.

Laufer, B. (1991). How much lexis is necessary for reading comprehension. In C. Lauren & M. Mordman (Eds.), Vocabulary and applied linguistics (pp. 126-132). Basingstoke: Macmillan.

Laufer, B., & Sim, D. D. (1985). Measuring and explaining the threshold needed for English for academic purpose texts. Foreign Language Annals, 18, 405-413.

Mezynski, K. (1983). Issues concerning the acquisition of knowledge: Effects of vocabulary training on reading comprehension. Review of Educational Research, 53, 253-279.

NAEP Reading Consensus Project. (n.d.) Reading framework for the National Assessment of Educational Progress: 1992-2000. Washington, DC: National Assessment Governing Board.

National Standards in Foreign Language Education Project. (1996). Standards for foreign language learning: Preparing for the 21st century. Lawrence, KS: Allen Press, Inc.

National Standards in Foreign Language Education Project. (1999). Standards for foreign language learning in the 21st century. Lawrence, KS: Allen Press, Inc.

Qian, D. D. (1998). Depth of vocabulary knowledge: Assessing its role in adults' reading comprehension in English as a second language. Unpublished doctoral dissertation. University of Toronto.

Swender, E. (Ed.). (1999). ACTFL Oral Proficiency Interview Tester Training Manual. Yonkers, NY: American Council on the Teaching of Foreign Languages.

Ulijin, J. M., & Strother, J. B. (1990). The effect of syntactic simplification on reading EST texts as L1 and L2. Journal of Research in Reading, 13, 38-54.

APPENDIX A. SAMPLE EXERCISES FOR SPANISH LANGUAGE ASSESSMENT

Sample exercises for each assessment area are presented on the following pages. No single task can illustrate all the facets of the task specifications or of the framework. These sample exercises are meant to be illustrative of the types of exercises that may appear on the Spanish NAEP. Although they embody specifications for the exercises, they do not, and cannot contain all specifications for exercises in any assessment area.

Interpretive Mode: Sample Listening-Based Exercise

For this assessment area, there will be multiple-choice and short answer exercises. Students will be asked to respond in English to demonstrate comprehension of a recorded text in Spanish. For example…

Imagine you are visiting Miami. Your Spanish-speaking friend Marta, who lives in Miami, has left a phone message on your answering machine regarding plans for tonight. Click on the button to hear the message.

When the button is clicked, the following message will be heard (but this text will not be seen).

Hola. Habla Marta. Te acuerdas que esta noche vamos a ver la última película con Antonio Banderas. Empieza a las once en punto. ¿Por qué no nos encontramos a comer a las siete y media en la Carreta en la calle ocho? Llámame. Estaré en casa toda la tarde. Hasta luego.

In order to focus their listening, students should read the question in English prior to hearing the audio.

1. What are tonight's plans, according to Marta?

Scoring Guide:

2	Student writes that you are going to meet (or eat) at 7:30 and that you are going to the movie at 11:00.
1	Student gives one piece of information
0	None of the above information is identified.

2. What does Marta want you to do when you get this message?

Scoring Guide:

1	Call Marta.
0	Another answer or no answer

Note: The above task reflects the *communities* goal because the setting is authentically in Miami and reflects the use of the language outside the classroom setting.

Interpretive Mode: Sample Reading-Based Exercise

For interpretive reading, there will also be multiple-choice and short answer exercises. Students will again be asked to follow directions and/or respond in English to demonstrate comprehension of a written text in Spanish for various reading purposes. For example…

Your friend Jorge wants to take some computer courses. He asks you to check ads and announcements on the Web for him. You see the following four ads. Which one would most interest him?

A. *Instituto de computación: busca a instructores de las aplicaciones comerciales de más demanda. Interesados llamen al 550-8861.*

* * * * * *

B. *Ventas: computadoras: Toshiba 4000 y Compaq Desktop en excelentes condiciones. Más detalles al tel. 526-2960.*

* * * * * *

C. *Es Usted: técnico, programador, consultante, etc...Anúnciese en este periódico. Haga de sus habilidades un negocio. 988-6245.*

* * * * * *

*D. *¡Visite nuestro centro hoy mismo! La matrícula para el verano empieza ahora. Se ofrecen seminarios en computación a todos niveles. Tel. 953-5692.*

*indicates correct option

Presentational Mode: Sample Writing-Based Exercise

For presentational writing, students will be asked to write in Spanish. Directions to the task will be given in English. For example, students may have 25 minutes to do the following…

Your school will host an exchange student from Mexico next year. You have been assigned to write a letter of introduction so that the exchange student will be familiar with your school. You may include information on where the school is located, the size, the teachers, extracurricular activities, rules, dress, school lunch, testing and grades, annual and daily schedule, course offerings and any other aspects you think important.

Be sure to write about any significant differences that the exchange student should know about.

Scoring Guide

Excellent Response • Communicates information clearly with an adequate number of well-chosen details • Expresses ideas with appropriate word choices • Is well organized, using transitions as appropriate • Exhibits variety in sentence length and structure • Contains few errors in grammar, spelling, and punctuation, which do not interfere with a Spanish speaker's comprehension of the writing • Demonstrates control of culturally authentic expressions and writing conventions in appropriate contexts • Includes information that shows knowledge of cultural differences and similarities
Sufficient Response • Communicates information with some details • May express ideas simplistically with some lack of variety in word choices • Is organized with ideas that are generally related, but has few or no transitions • Exhibits ability to write sentences, but sentence structure may be simple and unvaried • Shows some use of culturally appropriate writing conventions • Contains errors in grammar, spelling, and punctuation such that a Spanish speaker without a knowledge of English may have some difficulty comprehending the writing • Demonstrates some control of culturally authentic expressions and writing conventions • Includes information that shows some awareness of cultural differences and similarities
Uneven Response *(may be characterized by one or more of the following:)* • Communicates some information clearly, but may be like a list, undeveloped, or repetitive OR offer no more than a well-written beginning • Demonstrates inaccurate word choices, which interferes with expression of ideas • Is unevenly organized OR disjointed • Exhibits uneven ability to write sentences and little or no use of culturally appropriate writing conventions • Contains errors in grammar, spelling, and punctuation that make comprehension of the writing very difficult for any but those most used to reading such writing • Demonstrates little control of culturally authentic expressions and writing conventions • Includes little or no information showing awareness of cultural differences or similarities

Insufficient Response
(may be characterized by one or more of the following:)

- Presents fragmented information OR may be very repetitive OR may be very undeveloped such that no communication takes place
- Demonstrates inaccurate word choice OR inability to find words to express ideas
- Lacks organization so that thoughts are tenuously connected OR the response is too brief to detect organization
- Demonstrates minimal, if any, ability to write sentences
- Contains errors in grammar, spelling, punctuation, or usage (such as missing words or incorrect word use or word order) such that much of the response is incomprehensible
- Demonstrates no control of culturally authentic expressions and writing conventions

Interpersonal Mode: Sample Conversation-Based Exercises

Each exercise in this section of the assessment will have a total of five tasks. The first phase of the assessment—social conversation—consists of three tasks. The second phase consists of two role-play tasks. A student's performance on each task will be evaluated using a holistic rubric and will receive a single score. The five scores will be combined to produce one final score for each student on this assessment block. The communicative performance will be recorded on videotape for later scoring by trained raters.

Phase 1: Social Conversation

The contexts for the three tasks in this phase are, respectively, daily living, school, and work. For each task, the interviewer will draw from a standardized, limited inventory of initial and follow-up questions to conduct a brief conversation in which the student talks about himself or herself. Sample questions that could be used to initiate each task include the following:

Daily Living Task

- *Tell me about your daily routine during the week.*
- *What do you like to do in your free time?*

School Task

- *Tell me about your daily routine at school.*
- *Tell me something about your school.*
- *Tell me something about your favorite teacher.*

Work Task

- *Do you have a job? (to establish relevance) Tell me about it.*
- *What kinds of jobs do you and your friends have?*
- *Tell me about typical summer jobs for high school students around here.*

The development of each of the three tasks will take place through the use of follow-up questions, delivered by the interviewer as appropriate, drawn from a limited inventory of

questions. These follow-up questions are intended to enable students to speak to the best of their ability. Through them, students will produce an oral performance to be videotaped for later evaluation. The follow-up questions will serve the following purposes: simplification, expansion, elaboration, clarification. Some samples follow.

Simplification *(ask a simpler, more directed question)*

Question 1: *Tell me about your daily routine during the week.*
Follow-up 1 to Q1: *What time to you get to school?*
Follow-up 2 to Q1: *What do you do there?*

Expansion *(to elicit more detail, greater quantity, longer response)*

Tell me more.
Can you say more about this?
Can you describe _____for me?

Elaboration *(to introduce a new linguistic function at a higher level, such as compare/contrast, narrate in past time)*

Tell me about a time when_____.
Has _____ always been that way?

Clarification *(to elicit paraphrase)*

What does _____ mean?
Can you explain _____?

The oral performances produced by students will demonstrate their *comprehension* of the interviewer's questions and comments, the *comprehensibility* of their spoken Spanish in the context of the tasks, their ability to use *communication strategies* to enhance their comprehension and comprehensibility, and their *cultural awareness* and *cultural knowledge*. The interviewer's follow-up questions will provide students with the opportunity to demonstrate their abilities in these areas.

The choice of the interviewer's follow-up questions will be influenced by two factors: the *content* of the student's previous utterances, and the *performance level* of the student's previous utterances. Following are examples of how the daily living task might proceed differently for students at two levels of ability. The type of follow-up question is indicated in parentheses.

Example 1: With a low-level student producing just words or phrases.

Interviewer Question 1: What is a typical school day like for you?
Interviewer Question 2: What time do you arrive at school? *(Simplification)*
Interviewer Question 3: What do you do first? *(Expansion)*
Interviewer Question 4: And then? *(Expansion)*

Example 2: With a higher-ability student providing some detail and content.

Interviewer Question 1: What is a typical school day like for you?
Interviewer Question 2: What do you mean by (word or phrase used in response)? *(Clarification)*
Interviewer Question 3: Has (condition mentioned in response) always been that way? *(Elaboration)*
Interviewer Question 4: What happened at (event mentioned in response)? *(Expansion)*

Draft Sample Rubric for Phase 1 (Social Conversation) Conversation-Based Tasks

Score & Description
Excellent Performance • Demonstrates comprehension of both simple and elaborate questions • Provides detailed answers to questions • When asked to expand on a response, is able to provide additional details and to produce longer utterances • Can elaborate on a response when asked to do so • Shows accuracy and fluency when narrating or describing in present time; shows some accuracy if asked to elaborate in past time • Produces vocabulary from a variety of topic areas • Uses pronunciation and intonation patterns that are understandable to the interviewer • Successfully uses a variety of communication strategies to sustain the interaction, such as asking questions to get information and showing resourcefulness to express ideas for which vocabulary is lacking
Sufficient Performance • Demonstrates comprehension of simple questions, but comprehension may break down when attempting to understand more elaborate questions or questions not supported by the context • Provides simple personalized responses to questions or comments using sentences and short strings of sentences • When asked to expand on a response, is able to offer additional language only sporadically • Uses basic structures and simple sentences in the present tense; many inaccuracies occur when attempting more complex language • Produces vocabulary from familiar topic areas, though false cognates or English may be used when attempting to communicate beyond this scope • Uses pronunciation and intonation patterns that are understandable to an interviewer accustomed to interacting with language learners, though false starts and hesitancy may be frequent • Attempts to maintain the conversation by using communication strategies such as paraphrasing and question-asking

Insufficient Performance

- Demonstrates comprehension of high frequency vocabulary
- Relies on the interviewer's use of simplification strategies to comprehend the interviewer's questions and to produce own responses
- Demonstrates a reliance primarily on memorized words and phrases to answer highly predictable questions
- Produces vocabulary from a limited number of familiar topic areas
- Uses pronunciation and intonation patterns that are understandable to an interviewer accustomed to interacting with language learners, though false starts, pauses, and recourse to English may be frequent
- Primarily uses facial expressions and gestures to indicate comprehension problems and may clarify meaning, or express own meaning, by repeating words and phrases used by the interviewer

Unsatisfactory Performance

- No demonstration of comprehension of questions or statements
- Unable to answer questions
- Little or no language produced
- No attempt to use communication strategies to sustain the interaction

Phase 2: Role-Play Tasks

Note that during Phase 1, the student and the interviewer have not assumed roles; the student is talking about himself or herself, and the interviewer is an interested stranger who is trying to get to know the student better. In Phase 2, the two participants assume roles to carry out two tasks. The context for these tasks will be intercultural: the interviewer will be a representative from an organization, company, or school in the Spanish-speaking world, and the student will be either an individual with a personal interest in the task, or a representative of his or her school, town, organization, and so on. The task will require that the two participants work together to devise a plan, create a set of suggestions, or come to a decision. The interviewer will be supplied with the appropriate cultural information to play his or her role, and also will receive instructions on how to conclude the task.

Example of a role-play task: Making arrangements for a school-to-school exchange. The student will receive these instructions:

Your school is initiating a partner-school exchange program with a school in (name of city/country). You are having a meeting with a representative from the school to make suggestions for the clothing that students from each school will need to take with them on the exchange. Work together to find out what clothing teenagers wear in each community, noting similarities and differences. Take into account climate, school policies, and social customs.

The interviewer will receive these instructions:

You are a representative from a school in (name of city/country) that is establishing a partner-school exchange program with the student's school. You are having a meeting with the student to make suggestions for the clothing that students from each school will need to take with them on the exchange. Work together to find out what clothing teenagers wear in each community, noting similarities and differences. Take into account climate, school policies, and social customs.

The interviewer will also receive further instructions. Note the following:

1. The city and country will be predetermined and appropriate cultural information will be supplied.

2. The interviewer will also have a form that is formatted for both participants to take notes on. The interviewer will be trained to use the form to pace the task and to bring it to an end. The task should end when a list of suggestions of predetermined length has been compiled, or when a specific length of time has passed, whichever comes first.

Sample Form

Clothing Suggestions for Venezuela-US Exchange Program	
U.S. Students Should Pack	Venezuelan Students Should Pack
School:	School:
Social Activities with Friends:	Social Activities with Friends:
Formal:	Formal:

3. Although the instructions imply that there will be an equal exchange of information, the interviewer will have been trained to draw the student out so that the student will do most of the talking during the task and to encourage the student to ask questions to complete the task.

Draft Sample Rubric for Phase 2 (Role Play) Conversation-Based Tasks

Score and Description
Excellent Performance • Accomplishes the task with ease and facility; no major breakdowns in communication appear • Demonstrates comprehension of both simple and elaborate questions, as well as of more extended discourse, clarifying details by asking appropriate questions • Provides detailed answers to questions • Uses pronunciation and intonation patterns that are understandable to the interviewer • Shows accuracy and fluency when narrating and describing in present time; shows some accuracy when asked to elaborate in past time • Produces vocabulary from a variety of topic areas • When asked to expand on a response, is able to provide additional details and to produce longer utterances • Can elaborate on a response when asked to do so • Successfully uses a variety of communication strategies to sustain the interaction, such as asking questions to get information and showing resourcefulness to express ideas for which vocabulary is lacking • Demonstrates control of culturally authentic expressions in appropriate contexts • Includes information that shows knowledge of cultural differences and similarities, as appropriate • Asks knowledgeable questions to get cultural information when appropriate
Skillful Performance • Accomplishes the task with no major breakdowns in communication • Demonstrates comprehension of both simple and elaborate questions, clarifying details by asking appropriate questions • Provides personalized answers to questions, using strings of sentences and some details • Uses pronunciation and intonation patterns that are understandable to the interviewer • Uses basic structures and present tense with substantial accuracy; inaccuracies occur when asked to elaborate • Produces vocabulary from familiar topic areas • When asked to expand on a response, is able to offer additional details and produce longer utterances • When asked to elaborate, is rarely able to do so • Makes an effort to use communication strategies to sustain interaction, although not always successfully • Demonstrates some use of culturally authentic expressions in appropriate contexts • Includes some information that shows knowledge of cultural differences and similarities, as appropriate • Asks knowledgeable questions to get cultural information when appropriate

Sufficient Performance

- Accomplishes the task with support from interviewer when communication breaks down
- Demonstrates comprehension of both simple questions and simple discourse on familiar topics, but comprehension may break down when attempting to understand more elaborate messages or messages not supported by the context
- Provides simple personalized response to questions and comments using sentences and short strings of sentences
- Uses pronunciation and intonation patterns that are understandable to an interviewer accustomed to interacting with language learners, though false starts and hesitancy may be frequent
- Uses basic structures and simple sentences in the present tense; many inaccuracies occur when attempting more complex language
- Produces vocabulary from familiar topic areas, though false cognates or English may be used when attempting to communicate beyond this scope
- When asked to expand on a response, is able to offer additional language only sporadically
- Attempts to maintain the conversation by using communication strategies such as paraphrasing and question-asking
- May include some information that shows knowledge of cultural differences and similarities, as appropriate
- May ask simple questions to get cultural information when appropriate

Uneven Performance

- Accomplishes task with frequent support from the interviewer; some communication breakdowns are apparent
- Demonstrates comprehension of most simple questions and simple discourse on familiar topics; comprehension is not consistent
- Sometimes relies on the interviewer's use of simplification strategies to comprehend the interviewer's question or statements and to produce responses
- Provides simple responses to questions and comments, using a combination of memorized phrases and personalized sentences
- Uses pronunciation and intonation patterns that are understandable to an interviewer accustomed to interacting with language learners, though false starts and hesitancy may be frequent
- Uses basic structures and the present tense in simple sentences, although not consistently
- Produces vocabulary from familiar topic areas, though false cognates or English may be used when attempting to communicate beyond this scope
- Primarily uses facial expressions and gestures to indicate comprehension problems and may clarify meaning, or express own meaning, by repeating words and phrases used by the interviewer
- Provides limited evidence of knowledge of cultural differences and similarities

Insufficient Performance

- Accomplishes the task minimally with much support from the interviewer during frequent communication breakdowns
- Demonstrates comprehension of high frequency vocabulary
- Relies on the interviewer's use of simplification strategies to comprehend the interviewer's questions or statements and to produce own responses
- Demonstrates a reliance primarily on memorized words and phrases in highly predictable interactions
- Uses pronunciation and intonation patterns that are understandable to the interviewer, though false starts, pauses, and recourse to English may be frequent
- Produces vocabulary from a limited number of familiar topic areas
- Primarily uses facial expressions and gestures to indicate comprehension problems and may clarify meaning, or express own meaning, by repeating words and phrases used by the interviewer
- Limited or no ability to provide evidence of knowledge of cultural differences and similarities

Unsatisfactory Performance

- Unable to accomplish the task, despite help from the interviewer OR Unable to contribute to the accomplishment of the task
- No demonstration of comprehension of prompts, questions or statements
- Little or no language produced
- No attempt to use communication strategies to sustain the interaction
- No demonstration of cultural knowledge

APPENDIX B. PARTICIPANTS IN NAEP FOREIGN LANGUAGE CONSENSUS BUILDING PROCESS

The Consensus Steering Committee

Government

- Ray Clifford, Defense Language Institute, Monterey, CA
- Everette Jordan, U.S. Government Language School, Washington, DC

Business and Industry

- William A. Fleig, Prentice Hall, Glenview, IL
- Armando Guzmán, Univision, Washington, DC
- Helen Hamlyn, Language Testing International, White Plains, NY
- Jeffrey Munks, Arista Knowledge Systems, Alameda, CA

Parents

- Michelle Anciaux, Washington State PTA, Tacoma, WA

Private School Representative

- Sarah Donnelly, Association of Independent Maryland Schools (ret.), Glen Burnie, MD

Higher Education and Research

- Beverly Harris-Schenz, University of Pittsburgh, Pittsburgh, PA *(corresponding member)*
- G. Richard Tucker, Carnegie Mellon University, Pittsburgh, PA
- Guadalupe Valdés, Stanford University, Stanford, CA

Foreign Language and Policy Organizations

- Richard Brecht, National Foreign Language Center, Washington, DC
- Hiroko Kataoka, Japan Foundation, Santa Monica, CA
- Elizabeth Welles, Modern Language Association, New York, NY

State and District Foreign Language Supervisors

- Pat Barr-Harrison, Prince Georges County Public Schools, Upper Marlboro, MD
- Myriam (Mimi) Met, Montgomery County Public Schools, Rockville, MD
- Mario Nuñez, Florida Department of Education, Tallahassee, FL
- Pat Porter, Texas Education Agency, Austin, TX *(corresponding member)*

Foreign Language Consultant

- Protase Woodford, Educational Testing Service (ret.), Titusville, NJ

The Consensus Planning Committee

Teachers

- Donna Clementi, Appleton Public Schools, Appleton, WI
- Sandy Gutierrez, Fairfax County Public Schools, Falls Church, VA
- Deborah Lindsay, Greater Albany Public Schools, Albany, OR
- Maria Messina, Adrian C. Wilcox High School, Santa Clara, CA
- Jane Shuffleton, Brighton Central Schools, Rochester, NY

Higher Education Administration, Foreign Language Education and Research

- Andrew Cohen, University of Minnesota, Minneapolis, MN
- Antony Kunnan, California State University, Los Angeles, CA
- Judith Liskin-Gasparro, University of Iowa, Iowa City, IA
- Cecilia Pino, New Mexico State University, Las Cruces, NM
- June Phillips, Weber State University, Ogden, UT
- William Schafer (TAP liaison), University of Maryland, College Park, MD

Foreign Language Organizations and Consortia

- Jayne Abrate, American Association of Teachers of French, Carbondale, IL
- Sylvia Jones, California Foreign Language Project, Pasadena, CA
- Lynn Sandstedt, American Association of Teachers of Spanish and Portuguese, Greeley, CO
- Helene Zimmer-Loew, American Association of Teachers of German, Cherry Hill, NJ

State Foreign Language Supervisors

- Elizabeth Hoffman, Nebraska Department of Education, Omaha, NE
- Joan Patterson, Utah Department of Education, Salt Lake City, UT
- Paul Sandrock, Wisconsin Department of Education, Madison, WI

District Foreign Language Supervisors

- Martha Abbott, Fairfax County Public Schools, Fairfax, VA
- Yu-Lan Lin, Boston Public Schools, Boston, MA
- Martin Smith, Princeton Regional School District, Princeton, NJ

The Technical Advisory Panel

- Frances Butler, University of California, Los Angeles, CA
- Dan Eignor, Educational Testing Service, Princeton, NJ
- Kadriye Ercikan, University of British Columbia, Vancouver, BC
- Keiko Koda, Carnegie Mellon University, Pittsburgh, PA
- William Schafer (Planning Committee liaison), University of Maryland, College Park, MD
- Guillermo Solano-Flores, WestEd, Menlo Park, CA

The Project Management Team

The Center for Applied Linguistics (CAL)

- Dorry Kenyon, Project Director
- Nancy Rhodes, Associate Project Director
- Regla Armengol, Associate Project Director

The American Council on the Teaching of Foreign Languages (ACTFL)

- C. Edward Scebold, Networking Director
- Elvia Swender, Associate Networking Director

The American Institutes for Research (AIR)

- Beverly Farr, Technical Director
- Julia Mitchell, Assessment Specialist
- Steven Ferrara, Senior Technical Advisor
- Eugene Johnson, Senior Technical Advisor
- Archie Lapointe, Senior Technical Advisor