Court Interpreter Oral Examination:

Test Construction Manual

Consortium for Language Access in the Courts

Equal Justice: Bridging the Language Divide

The Mission of the Consortium is to inspire and enable its members to promote equal access to justice in courts and tribunals by eliminating language barriers for persons with limited English proficiency.

Copyright © 2010 by the National Center for State Courts, on behalf and for the use of the *Consortium for Language Access in the Courts*. All rights reserved.

The National Center for State Courts is an independent, nonprofit, tax-exempt organization in accordance with Section 501 (c)(3) of the Internal Revenue Code.

Except as permitted under the Copyright Act of 1976 and as otherwise expressly provided herein, no part of this publication may be reproduced in any form or by any means, electronic or mechanical, including the use of information storage and retrieval systems, without permission in writing from the copyright holder. Permission is hereby granted to Consortium members, Consortium staff and consultants to reproduce and distribute this publication for educational purposes if the copies credit the copyright holder.

For additional information, please contact:

Consortium for Language Access in the Courts
National Center for State Courts
300 Newport Avenue
Williamsburg, VA 23185

This document is also available on the website of the *Consortium for Language Access in the Courts*:
http://www.ncsc.org/Web%20Document%20Library/EC_StateInterpCert.aspx

Publication Date: April 2010

Court Interpreter Oral Examination:

Test Construction Manual

Consortium for Language Access in the Courts

The original version of this document was developed in 1996 under a grant from the State Justice Institute (SJI-95-12A-B164). Points of view and opinions stated in this report are those of the project staff and do not necessarily represent the official position or policies of the State Justice Institute or any member of the project advisory committee.

TABLE OF CONTENTS

Purpose of This Manual 1

[Section 1] **Overall Design of Tests** 1

[Section 2] **General Management of Test Development Activities** 3

[Section 3] **Compilation of Materials for Test Development and Production of Base Texts** . 4

[Section 4] **Qualifications of Test Development Teams** 6

[Section 5] **Test Development Process** 7

[Section 6] **Description of Scoring Units** 7

Table 1: Scoring Unit Descriptions and Testing Goals 8

[Section 7] **Selection and Classification of Scoring Units** 10

[Section 8] **Distribution of Scoring Units: Full 3-Section Examination** 12

Table 2: Scoring Unit Distribution for Standard Model 12

Table 3: Scoring Unit Distribution in the Consecutive 13

[Section 9] **Distribution of Scoring Units: Abbreviated Examination** 14

Table 4: Scoring Unit Distribution for Abbreviated Model 14

[Section 10] **Length of Test Sections** 15

Table 5: Number of Words in Test Sections 15

[Section 11] **Production of Test Materials for Test Administration** 15

[Section 12] **Production of Materials for Test Rating** 17

[Section 13] **Field Testing New Tests and Revising Existing Tests** 19

[Section 14] **Maintenance of Scoring Dictionaries** 21

[Section 15] **Documentation of the Test Development Process** 21

[Section 16] **Statistical Analysis of Tests** 22

[Appendix 1] **Sample Script** 23

[Appendix 2] **Sample Scoring Dictionary** 25

Purpose of This Manual

The Court Interpreter Oral Examination: Test Construction Manual was written by the Consortium for Language Access in the Courts' Technical Committee. The Technical Committee concerns itself with matters related to Consortium test development, test administration, test overviews and rating, statistics, and psychometric analyses. It also considers, drafts, and recommends modifications to Consortium manuals and to Sections 6 and 8 of the *Consortium Agreements*[1] which relate to Oral Test Instruments and Standards for Test Administration.

This Manual frames the Consortium's process of developing and maintaining valid and reliable oral performance examinations that members can use to credential persons who wish to become spoken language court interpreters, regardless of when or where they take the examination. The exams are developed to measure a candidate's ability to faithfully and accurately interpret the range of English ordinarily used in courtrooms into another language, and to understand and interpret into English what is said by a native speaker of another language.

The level of performance the test measures is the minimum acceptable level for entry into the profession of spoken language court interpretation. These exams do not measure other aspects of the knowledge, skills, and abilities one must have to perform the duties of a court interpreter, such as engaging in appropriate forms of situational control, dressing and conducting oneself in a manner consistent with the dignity of the court, and delivering services via telephone or as a member of a team.

This *Test Construction Manual* provides a structure for the oral examination development process to ensure that all tests have similar levels of difficulty, both for all languages and all versions within the same language.

[Section 1] Overall Design of Tests

Every test is developed according to specific test models and includes a scoring dictionary.

1.1 Test instruments shall include a section for one or more of the modes of court interpretation as follows:

- *Sight Translation:* the mode of interpreting whereby a written document in one language is interpreted into another language. This mode includes two separate components: one single-page document in English to be interpreted into another language and another single-page document in that other language to be

[1] See *Agreements for Consortium Organization and Operation,* Consortium for Language Access in the Courts, http://www.ncsc.org/Web%20Document%20Library/EC_StateInterpCert.aspx, Section 3.5.1.

interpreted into English. Using a written text that does not identify the scoring units, candidates read and deliver an interpretation of the written test instrument. A separate copy of each text that identifies the scoring units for the rating process is not provided to candidates.

- *Consecutive Interpretation:* the mode of interpreting performed when interpreting the testimony of a witness or when parties are in direct conversation with the Court on the record. Accordingly, this mode of interpretation is bidirectional. The test measures interpretation of some utterances from English into another language and other utterances from the other language into English. This section of the test is administered using a CD recording of a test text. The written test instrument has the scoring units marked and is not shown to the candidate, but is instead used during the rating process.
- *Simultaneous Interpretation:* the mode of interpreting that renders into another language everything that is said in English during court proceedings while the speaker continues to speak. This section of the test is administered using a CD recording of a test text. The written test instrument has the scoring units marked and is not shown to the candidate, but is instead used during the rating process.

1.2 Tests shall follow one of the two oral performance test models that the Consortium has approved as follows: [2]

- The "standard model" is the original model used since the formation of the Consortium and is sometimes referred to as a "full test" or a "whole test." It includes all three sections: sight translation, consecutive and simultaneous interpretation.
- The "abbreviated model" was adopted in 2003. This model includes a simultaneous section and a measure of conversational proficiency in English to be chosen by a member state from a list of available standardized tests promulgated and maintained by the Technical Committee.

1.3 Tests shall have a scoring dictionary. This is a reference document that the test development team prepares for test raters to use when scoring exams to ensure reliability in the rating process. It sets forth specific examples of how raters will determine whether certain renditions by candidates are to be counted as acceptable or unacceptable and is amended and updated from time to time according to established procedures (see Section 14). The following information will be included for each scoring unit on the test form: its number, classification, the actual word or words in the scoring unit, acceptable renderings, and unacceptable renderings (see sample in Appendix 2).

[2] The guidelines which are applied to determine when a test in a new language will be developed as a standard model or an abbreviated model can be found in Section 6 of the *Agreements for Consortium Organization and Operation*.

[Section 2] General Management of Test Development Activities

2.1 The Technical Committee is responsible for the development of new and modification of existing Consortium oral examinations and works closely with Consortium staff to ensure that test forms comply with the standards described in this Manual. Consortium staff may serve as the Test Development Manager (hereinafter "TDM"), or contract with a third-party consultant who meets the criteria required to serve in that capacity and who would operate under the general oversight of the Technical Committee and Consortium staff.

2.2 The TDM must have experience writing court interpreter examinations and possess an understanding of and follow the Consortium's policies for test development articulated in this Manual.

2.3 The TDM is responsible for assembling and supervising the test development team as outlined in Section 4, arranging for the professional review required per Section 3.5, and arranging for the linguistic review required under certain circumstances per Section 5.4. Consortium staff should solicit from Consortium members and other experts suggestions for:

- Members of test development teams and consultants hired to perform the reviews mentioned above; and
- Other forms of assistance, advice, and resources as Consortium staff deems appropriate.

2.4 Consortium staff shall develop, maintain, and revise as necessary, outlines describing the specific steps that the TDM, the test development team, and Consortium staff should follow when writing new tests. The outlines (one for the traditional test model and one for the abbreviated test model) will ensure that tests are developed in a consistent manner and that key steps in the process are not omitted.

2.5 The test development team shall conduct a preliminary assessment of the availability of bilingual legal glossaries of English and the corresponding test language. A list of such bilingual resources should be compiled by the team members and given to staff for distribution to Consortium members. If it is determined that there are no credible bilingual legal glossaries for the new test language, the team members will develop one as a component of the test development process.

2.6 The process should include a thorough written explanation of any special characteristics of the new test language that may present challenges for the provision of court interpreting services into or from that language, or that may require deviation from the test design described herein. This documentation should be kept in the official test file.

2.7 When a new examination has been developed, Consortium staff shall submit a report to the Technical Committee summarizing the development activities and describing any deviations from the standards set forth in this Manual and the reasons for such deviations.

[Section 3] Compilation of Materials for Test Development and Production of Base Texts

3.1 Consortium staff will compile a collection of transcripts of court proceedings and other documents that may serve as base texts for new exams. Members may submit such materials to staff for addition to this collection at any time. Staff will update the collection of resources as deemed appropriate by the Technical Committee.

3.2 The base texts which serve as the starting point for developing the parts of the sight translation section of an examination should be selected, if possible, from documents actually used by a court and originally written in the source language, such as:

- For English to other language sight translation
  - Police and other law enforcement reports
  - Investigative reports (e.g., pre-sentence report, child custody report) prepared for a court or a court support office
- For other language to English sight translation:
  - Correspondence to a judge and/or other court staff (e.g., character reference letters to a judge)
  - Documents prepared for or by a court to which parties are entitled access (e.g., stipulation, affidavit, court order, police report)

Documents that include text that is more or less established and non-variable should be avoided as they are readily available in the public domain and therefore could compromise test security. Examples include the Miranda warning, pre-printed forms, form letters, standard conditions of probation and other standard documents whose text does not vary.

The final text of the sight translation document in the other language shall be translated to English by the test development team and maintained in the test file.

3.3 The base texts which serve as the starting point for developing the consecutive and simultaneous sections of an examination should be selected from transcripts of official court proceedings that involve general legal concepts and procedures, and from types of legal situations that require large volumes of interpreting services. They should also represent a wide range of:

- Stages of legal proceedings (e.g., preliminary hearings, trials, sentencing).
- Types of proceedings and legal subjects heard in state courts (e.g., criminal, family, and civil courts).[3]
- Levels of proceedings (e.g., state, county, and, to a lesser degree, municipal courts).

3.4 The collection of base texts should not include materials that are:

- Unique to a particular state or a small subset of states, except where they can be easily adapted to meet Consortium standards.
- Specialized, arcane, or highly technical (e.g., reports or testimony of an expert witness, sophisticated legal argumentation).
- Replete with discourse that contains an inordinate number of legal terms (e.g., motions and legal arguments on technical points of law).
- Shocking in terms of content (e.g., containing graphic testimony in cases such as murder and sexual assault).

3.5 The documents should be reviewed and edited by Consortium staff to ensure that the discourse is clear, consistent and "tells a coherent story". These edits may include, but are not limited to, the deletion of repetitive utterances, false starts by speakers, and nonsensical words or phrases. In addition, the documents may be edited in the test development process as necessary to meet the requirements outlined in Sections 7 through 10 (see especially Section 7.14).

3.6 Every text edited by Consortium staff should be reviewed by professionals with the appropriate expertise to assess and confirm that the discourse of the text is consistent with acceptable legal practice and procedure.

3.7 The test writing team shall read and consider the scripts for content, cultural appropriateness and sensitivity. The TDM, in consultation with Consortium staff, shall select from the collection of reviewed base scripts the texts to be used by the test development team.

---

[3] Over time, the Consortium will increase the use of civil and family transcripts while maintaining a significant use of criminal transcripts.

Qualifications of Test Development Teams

4.1 Under the direct oversight of the TDM, each test should be written by a team consisting of at least two specialists as described below, one of whom must be a native speaker of the non-English language:

- One applied linguist, preferably a practicing professional interpreter who possesses the highest credentials available in the field; and
- One theoretical, scholarly linguist who has the most formal academic training possible in the linguistics of the language or, if such an individual is not available, the literature of the language, or, when no such trained linguist is available, a highly educated bilingual professional such as an attorney or teacher who is a native speaker of the language. Any such specialist selected to serve on the test development team will be trained to recognize the demanding specifics of court interpreting (e.g., maintaining the register of the source language; no editing, paraphrasing or embellishing).

4.2 When it is not possible to assemble professionals as described above, the team should consist of persons who come as close as possible to possessing these skills and credentials and who are deemed to be the most competent persons available for the purpose of writing a test.

4.3 If, prior to the development of an examination in a new language, there is no avenue through which a member of the test development team can become credentialed (certified, licensed, approved, etc.) and that member wishes to be so credentialed, he or she may notify the TDM of his or her desire to obtain the credential prior to commencing the test writing process. Once Consortium staff and the TDM have decided which simultaneous base script will be used, the TDM will administer the simultaneous section of the test and hold the recording until the new test has been completed.

Once the test is developed, the TDM will have the second member of the test development team rate the exam under his or her supervision. Any member of a test development team who passes the simultaneous test should be considered as having passed the test as if it had been taken under the usual circumstances and the TDM will recommend to that team member's home state (if that state is a member of the Consortium) that these test results be accepted for credentialing purposes.[4]

[4] This provision was adopted in view of the fact that, for many languages in which the Consortium will develop tests in the future, there is likely to be a very small pool of experts from which to select test developers and raters. It is essential to offer this option since, in its absence, prospective test developers and raters may decline to provide their services to the Consortium.

[Section 5] Test Development Process

5.1 The TDM shall develop a comprehensive training program that all new members of test development teams will complete that consists of the following:

- The TDM shall review the concept of scoring units as described in Section 6 with the test development team, and ensure that the team members have an understanding of how the examination, when completed, will be administered and scored.

- The TDM shall review this Manual, the *Test Writing Handbook*, and the test writing outline with team members prior to beginning the writing process.

5.2 The TDM shall guide the test development team in the production of scoring dictionaries.

5.3 When the test development team produces translations of any English text into another language, an independent linguistic review should be completed to ensure that the translated text is dialect-free and does not contain regional varieties that might not be recognized by all speakers of the language. The TDM shall explain to the test development team that they should produce a document that is a true dynamic equivalent of the English text and does not sound like a translation. The story, concept, or substance of the original text should be reproduced in the other language in a natural way that will be clear to the examinee.

5.4 When developing the consecutive section of the test, the TDM must understand and review with the test development team the meaning of the utterances that are in the non-English language to ensure that the substance of the questions and answers make sense and flow in a logical way. Team members shall provide the TDM with a final English version of the witness' answers.

5.5 A complete set of the test texts developed per Section 11 shall be retained for each test form identifying the number of words for each paragraph in the sight translation and simultaneous sections and for each utterance in the consecutive section.

[Section 6] Description of Scoring Units

6.1 Scoring units represent objective characteristics of language that the interpreter must understand and render appropriately during the interpretation. Each scoring unit is a word or phrase that captures a logically complete linguistic unit.[5]

[5] William E. Hewitt, *Court Interpretation: Model Guides for Policy and Practice in the State Courts,* National Center for State Courts, 1995, 104.

Tests are scored on the basis of scoring units, which are specific linguistic phenomena that interpreters must be able to deliver for a complete and accurate interpretation. A scoring unit is a pre-selected word or set of words in the exam material that is clearly identified by bolding and underlining the word or words comprising the scoring unit and then identifying the scoring unit in bolded superscript with a number and letter identifying the category of scoring unit to which it has been assigned. The following table lists the ten categories of scoring units and their corresponding testing goal(s).

Table 1: Scoring Unit Descriptions and Testing Goals

SCORING UNIT CATEGORY	DESCRIPTION	TESTING GOAL(S)
A: Grammar	"Grammar is a system of principles that govern the way a language works. Grammar describes how words relate to each other, particularly how they function in sentences."[6]	Ensure that candidates recognize and, within the limits of the source and target languages, satisfactorily handle the interpretation of grammar, especially verbs.
B: Language Interference	Terms or phrases that may invite misinterpretation due to interference of one language on another (e.g., false cognates, awkward phrasing, terms or phrases susceptible to literal renditions resulting in loss of precise meaning).	1-Measure the ability to keep languages separate, speaking them as an educated native speaker would, with no interference from the other language, and 2-Measure the ability to avoid being constricted unnecessarily by the source language resulting in interpretations that are literal or verbatim.
C: General Vocabulary	Vocabulary that is widely used in ordinary parlance and could be spoken by native speakers appearing in any courtroom.	1-Measure the ability to preserve lexical content of general source language terms when interpreted into the target language, 2-Measure the depth and range of candidate's vocabulary, and 3-Measure the ability to tap into a deep reservoir of vocabulary without hesitating or stumbling.
D: Legal Terms and Phrases	Any word or phrase of a legal or technical nature, or which is not common in everyday speech, but is commonly used in legal settings.	Measure the candidate's range of knowledge and recognition of common legal terms and styles of language used in courtrooms and the ability to faithfully interpret them into the target language, going into both languages, but especially from English into the other language.

[6] DiYanni, Robert, and Pat C. Hoy II. *The Scribner Handbook for Writers.* Boston: Allyn and Bacon, 1995, p. 221.

Scoring Unit Category	Description	Testing Goal(s)
E: Idioms and Sayings	An *idiom* is "a speech form or an expression of a given language that is peculiar to itself grammatically or cannot be understood from the individual meanings of its elements"[7]. *Sayings* are short expressions such as aphorisms and proverbs that are often repeated and familiar setting forth wisdom and truth.	Determine the candidate's breadth of knowledge and understanding of a language's common idioms and sayings, and the ability to interpret the meaning or an equivalent idiom or saying in the target language.
F: Register	Style of language drawn upon in various social settings; a key element in expressing degrees of formality, including curses, profanity, and taboo words. Register shows, through a pattern of vocabulary and grammar, what a speaker or writer is doing with language at a given moment.	Assess the candidate's ability to preserve the level of language so that others' impression of the speaker is not raised or lowered by the interpreter and assess the candidate's ability to interpret offensive terminology.
G: Numbers and Names	Any number, measurement, or proper name.	Measure the candidate's ability to be precise and accurate with all numbers, maintain weights and measures as stated in the source language without converting them to another system (e.g., from metric to English), preserve names of businesses, streets, etc. without interpreting them (except that "Avenue, "Street," etc. may or may not be interpreted, but the actual name is not to be interpreted), and conserve every letter of a spelled name in the order uttered.
H: Markers, Intensifiers, Emphasis and Precision	Any word or phrase giving emphasis or precision to a description (e.g., adverbs, adjectives) or statement (e.g., can be grammatical in form), including time (e.g., the day after tomorrow, last night, next week).	Ensure that the various ways of marking speech are preserved so the same degree of impact and precision is conveyed to the listener of the interpretation.
I: Embeddings and Position	Words or phrases that may be omitted due to position (at the beginning or middle of a long sentence, second in a string of adjectives or adverbs) or function (tag questions).	Ensure that candidates preserve all elements of the source language, especially those that they may deem to be "unimportant," or forget due to their location or function in the utterance.
J: Slang and Colloquialism	Slang and colloquialisms are informal, nonstandard words or phrases that are used in informal, ordinary conversation but not in formal speech or writing and are identified in standard dictionaries as "slang," "colloquialism," or "informal" or are listed in published dictionaries of slang and/or colloquialisms or in scholarly articles and books so identifying them. Slang items, which are coined by social groups, may be used in test texts only when they have passed into widespread usage across the United States.	Measure the candidate's range of knowledge of nonstandard, informal forms of speech and their ability to interpret the meaning of such words and phrases without being bound to preserve their low register.

[7] *The American Heritage College Dictionary*, Third Edition, Boston: Houghton Mifflin Co., 1997, p. 674.

[Section 7] Selection and Classification of Scoring Units

The TDM will guide the test development team in understanding and abiding by the following principles of assigning scoring units during the course of developing a new test, when reviewing existing tests during the field testing stage, or as part of ongoing assessments of existing tests. The TDM and the team members should also closely examine and use the *Test Writing Handbook* as a guide.

7.1 The test development team, with oversight of the TDM, should select scoring units using the grids provided in the *Test Writing Handbook* as an aid. Specific types of scoring units should follow, as much as possible, the systematic structure provided therein.

7.2 A scoring unit may be any word or set of words in the base text regardless of their perceived importance in the discourse. It does not have to be a critical or important part of the discourse.

7.3 The discourse in which a scoring unit is embedded shall provide a clear context so the meaning of the scoring unit is unambiguous.

7.4 The ways the scoring unit can be correctly rendered in the target language should be clear and unambiguous.

7.5 The test development team should remember that there is often, if not usually, more than one right way to interpret a word or phrase and there are always many ways to incorrectly interpret a word or phrase.

7.6 Words and phrases being considered for scoring units in English source materials should be used in contemporary parlance by a wide range of native English speakers in the United States and its possessions, including lawyers, judges, and witnesses, and may not be obscure, archaic, outdated, or unique to a particular region, dialect, or other subset of speakers.

7.7 Words and phrases being considered for scoring units in a source language other than English should be used in contemporary parlance by a wide range of its native speakers in all countries where the language is spoken, including witnesses and parties, and may not be obscure, archaic, outdated, or unique to a particular country, region within a country, or other subset of speakers of that language.

7.8 Loan words and other types of interference in a mother tongue that arise due to influence of the English language should not be included in the test materials, but if they are, they should never be selected as scoring units.

7.9 The same or nearly same text should not be used as a scoring unit more than once in an examination. In addition, a particular verb form (i.e., tense, mood) should appear as a grammatical scoring unit no more than once in a test section.

7.10 The number of legal terms that are not scoring units should not exceed the number of legal terms that are scoring units in the final text of any test.

7.11 Care should be taken to distribute scoring units so they appear throughout the test as described in Sections 8 and 9, with adequate space between them to facilitate the rating process and to ensure that they do not appear in clusters at the beginning, in the middle, or at the end of a test section.

7.12 A scoring unit should be clearly associated with one of the ten categories of scoring units defined in Section 6. Any word or phrase in a base text could be selected as a scoring unit and many could likely be classified in more than one of the categories of scoring units. When a word or phrase is under consideration to be selected as a scoring unit, and the test development team is not sure how to classify the item, it should consider the guidance provided in the *Test Writing Handbook*. If there is still confusion or disagreement, the members should either avoid making the item a scoring unit or make a clear decision as to how to classify it and note the reasons why it has been classified as such in the scoring dictionary.

7.13 A scoring unit should consist of as few words as linguistically appropriate, rarely exceeding four words. The test development team should avoid phrases that include enough words to confuse or diffuse the testing goal as outlined in Table 1 or that can be reasonably viewed to include two different types of scoring units.

For example, *"the tall gentleman,"* which includes very few words and might be intended as a single scoring unit, is not an acceptable scoring unit because it includes multiple types of scoring units, among them the following examples:

- *Grammar* to measure article/noun or adjective/noun agreement of number and/or gender;
- *Markers and intensifiers* to assess accuracy of interpretation of the adjective "tall;"
- *General vocabulary* to test range of vocabulary with respect to "gentleman" and perhaps even "tall;"
- *Register* to ensure that "gentleman" is not lowered in register to "man," "guy," "fellow," or another similar term.

7.14 Suggested steps for identifying scoring units:

Step One: Read through a base text and preliminarily assign scoring units.

Step Two: Using the parameters set forth in Section 8 (for a full 3-section examination) or in Section 9 (for an abbreviated examination), calculate the types of scoring units that are generated in step one. Identify those which are over-represented and those which are under-represented in each test section.

Step Three: For those that are over-represented, select the units that appear best to keep (based on perceived strength of the scoring units themselves as well as on their

proximity to other scoring units) and eliminate as many as may be necessary to approximate the target number for each type.

Step Four: For those that are under-represented, it may be necessary to edit the text in such a fashion as to add material that includes these types of scoring units. The test development team should be careful when adding text to keep the total number of words within the limits allowed in Table 5.

[Section 8] Distribution of Scoring Units: Full Three-Section Examination

8.1 When a full test is being developed, the scoring units should be distributed among the test sections so that each test follows the presumptive distribution outlined in Table 2. The test development team should review the appropriateness of these percentages for each language combination and recommend adjustments to them when the linguistic properties of English and the other language make an adjustment necessary. It is acceptable for the TDM to approve deviation from the target indicated in the "Target %" column in the table below by plus or minus 10 percent, but if the deviation is significant, the TDM should consult with members of the Technical Committee for input and direction.

Table 2: Scoring Unit Distribution for Standard Model

Scoring Unit Category	Sight - English to Other Language	Sight - Other Language to English	Consecutive	Simultaneous	Unit Total	Target %
A: Grammar	4	4	15	10	33	15
B: Language Interference	3	3	9	6	21	10
C: General Vocabulary	8	8	15	13	44	20
D: Legal Terms and Phrases	3	3	11	16	33	15
E: Idioms and Sayings	0	0	7	4	11	5
F: Register	1	1	5	3	10	5
G: Numbers and Names	1	2	6	5	14	7
H: Markers, Intensifiers, Emphasis and Precision	3	3	9	7	22	10
I: Embeddings and Position	1	1	9	8	19	9
J: Slang and Colloquialisms	1	0	4	3	8	4
Total Number of Units	**25**	**25**	**90***	**75**	**215**	**100**

* Some early versions of the Consortium's exams included 75 scoring units in the consecutive section. Subsequent exams include 90 scoring units.

8.2 Distribution of Scoring Units in the Consecutive Section

When drafting the consecutive section of an examination, the utterances to be interpreted should meet the following standards:

- They should be of varied lengths, ranging from one word to no more than fifty words.
- When drafting the consecutive section, the test development team shall abide by the following provisions at the beginning of the test text in order to give examinees a chance to get started and into the rhythm of consecutive interpretation:
  - No more than twenty words and no scoring units should be used in the first two utterances.
  - In the next two utterances, the number of scoring units should be relatively small, and whatever scoring units are included should not be especially difficult or challenging.
- The scoring units must be embedded in utterances that vary in length approximating the distribution shown in Table 3 below.
- As illustrated in Table 3, the percentage of scoring units found in utterances ranging from 21 to 30 words must always be greater than the percentage of scoring units found in utterances ranging from 11 to 20 or 31 to 40 words.
- The percentage of scoring units in utterances ranging from 11 to 40 words may deviate by up to 10 percent from the percentages indicated in the table below.

Table 3: Scoring Unit Distribution in the Consecutive

Utterance Length (*In Number of Words*)	**Target %** English to Other Language	**Target %** Other Language to English
1-10	10%	10%
11-20	25%	25%
21-30	30%	30%
31-40	25%	25%
41-50	10%	10%
Total %	**100%**	**100%**
Number of Scoring Units	**40**	**50**

8.3 Distribution of Scoring Units in the Beginning of the Simultaneous Section

When drafting the simultaneous section of any examination, no scoring units should be included in the first ten seconds of the recording (i.e., approximately the first twenty words) in order to give the examinee a chance to get started.

Distribution of Scoring Units: Abbreviated Examination

9.1 When an abbreviated test is being developed, the scoring units should be distributed within the test as outlined in Table 4. The test development team should review the appropriateness of these percentages for each language combination and recommend adjustments to the TDM when the linguistic properties of English and the other language make an adjustment necessary. It is acceptable for the TDM to approve deviation from the target number of scoring units by plus or minus 10 percent, but if the deviation is significant, the TDM should consult with members of the Technical Committee for input and direction.

Table 4: Scoring Unit Distribution for Abbreviated Model

Scoring Unit Category	Total # of Units	Target % of Units
A: Grammar	16	21%
B: Language Interference	0	0
C: General Vocabulary	14	19%
D: Legal Terms and Phrases	13	17%
E: Idioms and Sayings	4	5%
F: Register	4	5%
G: Numbers and Names	5	7%
H: Markers, Intensifiers, Emphasis and Precision	8	11%
I: Embedding and Position	8	11%
J: Slang and Colloquialisms	3	4%
Total Number of Units	**75**	**100**

9.2 The length and assignment of scoring units at the beginning of the simultaneous section in the abbreviated model shall comply with the requirements for the same test section in the full three-section examination and the words should be counted in the same manner as described in Section 10.2.

[Section 10] Length of Test Sections

10.1 The number of words in each test section must be within the ranges shown in Table 5.

Table 5: Number of Words in Test Sections

Section of Test	Range in Number of Words
Sight translation English → Other Language Other Language → English	400-450 200-225 200-225
Consecutive English → Other Language Other Language → English	850-950 400-450 450-500
Simultaneous (for both standard and abbreviated models)	800-850
Total	**2050-2250**

10.2 Words in the sections of an examination should be counted as follows:

- Any word that would be separated from other words by a space or punctuation when written constitutes one word.
- Each portion of a hyphenated word counts as one word (e.g., "twenty-three" counts as two words).
- Numbers should be counted the way they would be written out as words (e.g., "2006" would be "Two thousand six" and count as three words) or spoken aloud (e.g.," Two thousand six," or "Two thousand and six").
- The words being counted are words in the source language, not the target language.

[Section 11] Production of Test Materials for Test Administration

The TDM, test development team, and Consortium staff are responsible for producing test texts that are ready for administration by test proctors. This section describes how the text for each section of an examination should be prepared for that purpose.

11.1 The sight translation documents to be handed to the candidates during the first section of the examination should be produced as follows:

- Both parts should include only the text of the test material and not include any headers, footers, or other writing that is not test content. The margins of the document should be set at one inch at the top, bottom, left, and right of an 8.5 x 11 standard white sheet with no watermarks.
- The text should fit on a single sheet of paper. In the event that the text is too long to fit on a single sheet, the font size should be adjusted as little as is necessary to fit the text on one sheet.
- No scoring units should be identified by bolding, underlining, or including identifiers in superscript.
- The text should be arranged in paragraph format, with the first line of each paragraph indented.
- The document written in English should be typed in Times New Roman font, size 12, double-spaced, and left-justified in portrait layout.
- The document written in the other language should follow the rules for the language's writing system per the leading authority for writing the language, and follow the same font requirements as mentioned above to the degree that they apply to the orthography of the language.
- The document written in the other language should follow the appropriate orthographic rules of the language, including the production of more than one version if the language has more than one written form (for example, traditional and simplified for Cantonese and Mandarin).

11.2 Production of CD recordings for the consecutive and simultaneous sections of the examinations:

- The recording for both the consecutive and simultaneous sections should be produced as follows:
  - The speakers should strive to emulate the quality of broadcast voice.
  - The recordings should be produced so that they are clear, free from extraneous noise (including hiss), and easily understood.
- The recording by which the consecutive section is administered should be produced as follows:
  - It should be recorded at a conversational rate of speech, neither fast nor slow, clearly enunciated, and with enough expression to appropriately emote the text.
  - The person who reads the English portion of the text should be a native speaker of English and the person who reads the non-English portion of the text should be a native speaker of that language.

- A standardized amount of pause time between tracks should be inserted by appropriate software so that individuals proctoring the exams expect the same amount of pause time and can seamlessly and consistently administer the consecutive test regardless of the language.
- The gender of the person reading each role should match the gender of the role he or she is reading.

11.3 The recording by which any simultaneous section is administered should be produced as follows:

- It should include a standardized, 45- to 60-second introduction, which includes identification for the examinee of the first sentence or phrase of the test text followed by five seconds of silence before the test begins.
- It should be recorded throughout at a speed of 120 words per minute, as far as is possible. No single minute should be slower than 110 words per minute or faster than 130 words per minute. The introduction should be recorded at the same volume as the test content.
- The following statement should be recorded after a 10-second pause at the end of the simultaneous test so that it is clearly not part of the test: "The test is now over; please remove your headphones."

[Section 12] Production of Materials for Test Rating

12.1 The final versions of the test texts should conform to the following editorial guidelines, which are illustrated in Appendix 1:

- Margins should be set at 1 inch at the top, bottom, left, and right of an 8.5 x 11 standard white sheet, with no watermarks.
- Document text should be in Times New Roman font, font size 12, double-spaced, and left-justified in portrait layout.
- All scoring units should be bolded and underlined, and each scoring unit should be identified in superscript by a number and type identification, which should be bolded, but not underlined. For example: **construction manual**[1C]
- Sometimes there are scoring units that consist of more than word and are not all contiguous to each other on the same line. There are two instances when this can occur:

When the words of a scoring unit appear at the end of one line and the beginning of the next line, the bolded number and type of scoring unit shall be provided in superscript at the end of the last word of the scoring unit as follows:

> "We expect to hear that witness testify **the day after tomorrow** [29H]."

- When a word or words that are not part of a scoring unit are located between words that constitute a scoring unit, the text should be marked as follows: "The suspect clearly intended to steal **not only** the car stereo **but also**[12A] the GPS."

- Sometimes punctuation precedes or immediately follows a scoring unit. When this happens, the following guidelines shall be followed in producing the test text:
  - Punctuation markers (e.g., comma, period, question mark, exclamation point, open or close quotation mark) should not themselves be bolded or underlined.
  - When there is a punctuation marker immediately following the last letter of a word in a scoring unit, the superscripted identifying information should be inserted before that punctuation marker, as illustrated in the following example: "…his eyes rolled back in his head and he **collapsed**[25C]."
- Each script shall be identified in the header of the document, in Times New Roman font, bolded, font size 14. The header should contain:
  - at the left margin, the test form identification,
  - in the center, identification of the test section and part, if applicable, and
  - at the right margin, the script identification.
- At the bottom right of the last page of each test section and the two parts of the sight translation section, there should be a grid for raters to complete, with the summary of the point score for that test section. See Appendix 1 for an example.

12.2 The scoring dictionaries that are created for each test section should adhere to the following editorial guidelines, which are illustrated in Appendix 2:

- The scoring dictionary should be created as an Excel document in landscape layout.
- The header should be typed in Times New Roman font, bolded, font size 14, and should identify the Language, Test Form, and Script Identification.
- Column headings should be as follows:
  - Column 1: No.
    - This column should contain the number of the scoring unit within the test sections (i.e., 1 through 25, or 1 through 90, etc.)

- Column 2: Scoring Unit Type and Description
  - This column should contain the alpha identification of the type of scoring unit (A, B, C, D, E, F, G, H, I, or J), a colon, the name of the scoring unit category (e.g., Legal Terms and Phrases or Grammar).
- Column 3: Scoring Units
  - The word(s) or numbers of the actual scoring unit should be provided in this column.
- Column 4: Acceptable Response(s)
  - This column should contain all of the acceptable interpretations of the scoring unit, separated by semi-colons.
- Column 5: Unacceptable Response(s)
  - This column should contain all of the unacceptable interpretations of the scoring unit, separated by semi-colons.
- Column 6: Notes
  - This column should contain any instructions, hints, or explanations for test raters about why the test development team classified a particular scoring unit as they did or other guidance on how to rate it. This column will later contain additional notes from the raters.

[Section 13] Field Testing New Tests and Revising Existing Tests

13.1 Some form of field testing shall be conducted on every new test before it is released to members for routine administration and rating. Relying on the advice and expertise of the specialists serving on the test development team and any other expert the TDM selects, the TDM will coordinate the form of field testing listed below that is most appropriate for a given language. Field testing is a simulation of operational test administration that identifies scoring units that do not function as intended or pose other problems for candidates or raters, identifies additional acceptable and unacceptable responses for the scoring dictionary, and uncovers other problems in the structure, content or mechanics of a particular test. After field testing, the TDM and specialists must make needed adjustments identified in the field testing to the test instrument before it is considered an operational product. The options for this phase of test development are as follows:

- For languages which are spoken by two or more distinct speech communities around the world and for which there is a large number of prospective candidates, field testing should be administered to a sample of test takers who may or may not be

taking the test in order to obtain a court interpreting credential, come from at least two geographically separate parts of the United States, and represent as many different speech communities of the language as possible. Any of these components may be relaxed if the TDM finds that it is not feasible and documents the reasons for reaching that conclusion.

- For languages which are spoken by two or more distinct speech communities around the world and for which there is a small number of prospective candidates, field testing should ideally be administered to a sample of test takers who may or may not be taking the test in order to obtain a court interpreting credential and represent at least the language's largest speech communities represented by residents of the United States. If the sample includes test takers who seek a court interpreting credential through the field test, the results shall be considered final only after all field test candidates have been tested, all field tests have been analyzed, and the scoring units and other necessary adjustments have been made, and the field tests are then graded on the basis of the final text(s) of the test instrument(s).
- For languages which are spoken by essentially a single speech community, field testing should ideally be administered to a sample of test takers who may or may not be taking the test in order to obtain a court interpreting credential. If the sample includes test takers who seek a court interpreting credential through the field test, the results shall be considered final only after all field test candidates have been tested, all field tests have been analyzed, and the scoring units and other necessary adjustments have been made, and the field tests are then graded on the basis of the final text(s) of the test instrument(s).
- When a test is being developed in any language for which the foregoing options are not feasible, field testing must be given to a minimum of two candidates and the results shall be considered final only after all field test candidates have been tested, those field tests have been analyzed, and the scoring units and other necessary adjustments have been made, and the field tests are then graded on the basis of the final text(s) of the test instrument(s).

13.2 As long as any test form is being administered, raters who grade the test will be engaging in ongoing review and evaluation of that test. For example, test raters sometimes discover that scoring units do not perform well over time or identify shortcomings in the selection and/or classification of scoring units and they will provide such feedback to the TDM. The TDM will consult with the specialists working on that test development team and issue whatever revised versions are warranted. When there is a sufficient number of revisions indicating that a given test has undergone considerable change, it may be retired, revised, and reissued under a new number (e.g., 1.2 would be the second version of test #1, 1.3 would be the third version of test #1).

[Section 14] Maintenance of Scoring Dictionaries

14.1 Consortium staff will designate two experts to serve as a team for maintaining the scoring dictionary for each language for which tests are developed. That list will be made available to the Technical Committee in whatever manner the committee indicates and may be revised from time to time, as determined necessary or desirable by Consortium staff and approved by the committee.

14.2 When the raters are rating examinations, they will hear renditions that do not appear in the scoring dictionary for that examination. If, in their opinion, the rendition is one that candidates may use in the future (either a new way of accurately interpreting a scoring unit or incorrectly interpreting a scoring unit), they must complete and submit a Scoring Unit Suggestion Form approved by the Technical Committee according to procedures set forth in the Test Rating Manual.

14.3 Consortium staff will forward the suggestion to two designated experts, who will research the suggestion and inform staff of their agreement or disagreement with the suggestion and staff will update the scoring dictionary and notify the rater who made the suggestion. If Scoring Unit Suggestion Forms are received from raters who are rating an examination that is written in an alphabet other than the Roman alphabet, staff will contract with an original member of the test development team or a designated, experienced test rater to update the scoring dictionary and return it to staff in PDF format.

14.4 If the expert(s) reviewing the suggested revisions disagree with the suggestion or are uncertain of the accuracy of the suggestion, they may confer with other raters or members of the test development team before notifying Consortium staff of their agreement or disagreement with the suggestion.

[Section 15] Documentation of the Test Development Process

15.1 Consortium staff should collect and preserve information about the development and maintenance of each test, including the names and qualifications of every person involved in the writing and review of the test instrument.

15.2 The TDM will document for staff any deviations from the construction standards as described in this Manual, articulating why such deviation was necessary, and noting any specific challenges or problems faced by the test development team and reviewers. Consortium staff will keep the Technical Committee informed of changes and revisions made to test instruments and deviations from the standards herein.

Statistical Analysis of Tests

16.1 After an adequate sample of examinees has taken a given test version, Consortium staff should conduct an analysis of the test to address any reliability and validity issues that may arise. Information obtained through the analysis will be used to revise or delete poor performing items from future test versions. Staff will report the findings of the analysis and make recommendations to the Technical Committee and modifications will be made accordingly.

Appendices

Good morning. My name is Janet Smith and I am an assistant **state's attorney**[1D] in our **county**[2C]. Ladies and gentlemen of the **jury**[3D], we believe that the evidence will show that during the morning **hours**[4I] on **July 2nd, 2004**[5G], a woman by the name of **June Jones**[6G] was brutally **beaten**[7A] and assaulted in the living room of her home here in the city of Madison. She was held down on her living room floor for two hours, and her **assailant**[8B] took her **purse**[9C] when he **finally**[10I] ran away. **We will prove**[11A] that the **defendant**[12D], **Omar Butler**[13G], is the person who assaulted, restrained and **stole from**[14A] June Jones. The State has **charged**[15D] Mr. Butler with **attempted sexual assault**[16D], **false imprisonment**[17D] and theft.

June is a 42 year old **divorcée**[18C] who lives **alone**[19I] at **1729 Rosewood Avenue**[20G]. On the night of **Friday, July 1st** [21G], June was at her house with a friend, **hanging out**[22J] and drinking. **Late**[23I] in the evening, the victim told her friend that she wanted to go over to Harry's Bar. Her friend said that he wanted to **call it a night**[24E], so June decided to go by herself.

Possible Points	
(-) # Incorrect	
Total Correct	

[Appendix 2]

Sample Scoring Dictionary

No.	Scoring Unit Category	Scoring Units	Acceptable Response(s)	Unacceptable Response(s)	Notes
1	D: Legal Terms/Phrases	state's attorney	*Acceptable*	*Unacceptable*	*Notes from members of the test development team to test*
2	C: General Vocabulary	county	*interpretations*	*interpretations*	*raters (or notes by test raters*
3	D: Legal Terms/Phrases	jury	*of the various*	*of the various*	*for the benefit of subsequent*
4	I: Embeddings/Position	hours	*scoring units would*	*scoring units would*	*raters) would be entered in*
5	G: Numbers & Names	July 2nd, 2004	*be entered in these*	*be entered in these*	*these spaces.*
6	G: Numbers & Names	June Jones	*spaces.*	*spaces.*	
7	A: Grammar	beaten			
8	B: Language Interference	Assailant			
9	C: General Vocabulary	purse			