



Scoring the Consortium for State Court's Oral Proficiency Exam for Interpreters

By Wanda Romberger

The Consortium for State Court Interpreter Certification (Consortium) is a multi-state partnership consisting of 40 states that have come together, pooling financial resources and professional expertise to develop and share oral proficiency examinations to measure the abilities of court interpreters.¹ Since 1995, the Consortium has developed 21 oral examinations in 16 languages. Interpreter program managers in member states use the exam results to identify interpreters who are qualified to work in the state courts, as well as those who are close to satisfying the requirements. States can then decide to credential qualified interpreters and whether or not to invest scarce financial resources to provide appropriate training for candidates who do not yet qualify but show promise.

Minnesota, New Jersey, Oregon, and Washington established the

Exam raters must strive for complete neutrality, fairness, and objectivity.

Consortium in July 1995. This came about as the result of research conducted by the National Center for State Courts (NCSC) between 1992 and 1995 to examine the nature and scope of management problems related to interpreter services in the state courts. At that time, there was still inconsistency among the interpreting programs offered by the founding states. Minnesota and Oregon had pre-existing oral examinations developed in-state, while New Jersey and Washington had the finances to create a pool of resources

for additional test development. The study by NCSC indicated that establishing an interstate authority with the capacity to coordinate test development efforts and share financial resources on a national scale was both desirable and feasible.²

Committed to improving their interpreter programs in their states and with the support of NCSC, representatives from the Consortium's founding states met with William E. Hewitt, the principal court research consultant at NCSC. Representatives discussed, debated, and compromised

until, finally, detailed test development standards were articulated and documented. The standards called for a three-part exam with four segments, consisting of two sight translation exercises (English into the other language, and from the other language into English), a consecutive interpreting test, and a simultaneous interpreting test.³ This structure has remained essentially the same since the first exam was developed.

Once appropriate test development standards were agreed upon, Consortium representatives turned their attention to formulating standards for administering and rating (scoring) the examinations. The Consortium now maintains standardized manuals for test construction, test administration (including a candidate information booklet), and test rater training. To become a member of the Consortium, states must agree to administer and rate the organization's examinations in accordance with these standards. The following sections will give a general overview of the standards for test development and rating. For the complete official agreements for Consortium organization and operation, please visit www.ncsconline.org/D_Research/CourtInterp/Agreements2008FinalMay.pdf.

Developing the Consortium Oral Exam

In order to understand the Consortium's exam rating system, it is necessary to describe how the exams are developed. The first step in developing an oral examination is to gather and review transcripts of actual state court proceedings and formal court documents for appropriateness of content, language, and case type. When transcripts are identified that lend themselves well to testing, they are edited to

change the names of the parties involved in the original proceedings and to ensure that the material includes enough robust and varied linguistic challenges to meet the requirements of the Consortium's test construction manual. Segments from these transcripts and court documents are then selected for the two sight translation exercises and the consecutive and simultaneous interpreting tests.

Scoring Units

The next step in the test development process is for the test writers to identify appropriate scoring units within the text.⁴ Scoring units are special linguistic characteristics that interpreters must be able to render in order to deliver a complete and accurate interpretation. Each exam should contain 215 scoring units distributed among its four segments. There are 10 types of scoring units, including:

- **Grammar and Usage Types:**
 - grammar/verbs
 - false cognates/language interference
- **General Lexical Range Types:**
 - general vocabulary
 - legal terminology
 - idioms/sayings
- **Conservation Types:**
 - register
 - numbers/names
 - markers/intensifiers/emphasis/precision
 - embeddings/position
 - slang/colloquialisms

Each type of scoring unit is thoroughly explained to test writers, and they receive instructions to help them identify appropriate words and phrases to be used as scoring units. For example:

- For “embeddings/position” type scoring units, test writers are instructed to identify words or phrases within the text that are likely to be omitted due to their position within the utterance or their function.
- For “register” type units, test writers are instructed to identify words and phrases of unquestionably high or low register that can be preserved in the target language, but that might be lowered or raised incorrectly by unqualified or unprepared test candidates.

Consortium staff at NCSC work closely with the test writers. Because the 215 scoring units must be located, classified, and distributed, a full three-part oral examination can take up to 10 working days to develop. The staff documents any particular challenges encountered due to language constructs or limitations. The 215 scoring units are carefully distributed throughout the examination text in accordance with the scoring unit distribution standards found in the Consortium's test construction manual. Additional test construction standards dictate the number and types of units that must be found in each of the four segments of the examination. (See Table 1 on page 26 for a better understanding of the distribution of the scoring units.)

Scoring Dictionaries

The final part of the test writers' responsibility is to create a “scoring dictionary” for each segment of the examination. The scoring dictionary lists each scoring unit and its classification, complete with interpretations for each unit that have been marked “acceptable” or “unacceptable” by other teams of raters in the past. ➡

Table 1: Distribution of Scoring Units

Type of Scoring Unit	Sight Translation Exam: Other Language into English	Sight Translation Exam: English into Other Language	Consecutive Interpreting Exam	Simultaneous Interpreting Exam	Overall	Overall %
A. Grammar/Verbs	4	4	15	10	33	15
B. Cognates/Interference	3	3	9	6	21	10
Subtotal	7	7	24	16	54	25
C. General Vocabulary	8	8	15	13	44	20
D. Legal Terminology	3	3	11	16	33	15
E. Idioms/Sayings	0	0	7	4	11	5
Subtotal	11	11	33	33	88	40
F. Register	1	1	5	3	10	5
G. Numbers/Names	1	2	6	5	14	7
H. Markers, etc.	3	3	9	7	22	10
I. Embeddings, etc.	1	1	9	8	19	9
J. Slang/Colloquial	1	0	4	3	8	4
Subtotal	7	7	33	26	73	35
TOTAL	25	25	90	75	215	100

Trainers move through the mechanics of scoring with the new raters, sharing tips for efficient and accurate scoring and explaining how to complete the results report form.

In many cases, depending upon the language and the content of the scoring unit, there may be more than one perfectly acceptable rendition, especially if the language includes regional varieties of meaning. Test writers consult only reputable published bilingual dictionaries, idiom dictionaries, legal glossaries, and many other resources to

discover acceptable interpretations.⁵ Table 2 on page 27 shows an example of what a section of a scoring dictionary might look like and how it can be used. (Note that because it is all in English and includes no actual scoring unit text, Table 2 is not completely accurate. It is only meant to serve as an example.)

Scoring dictionaries are dynamic

documents to which test raters (sometimes referred to as scorers or examiners) are expected to update on a regular basis. When a test candidate interprets a scoring unit in a way that is not already included in the scoring dictionary, raters conduct the necessary research (always keeping the context of the scoring unit in mind) and add the interpretation to the scoring dictionary, either as an acceptable interpretation or an unacceptable one. Before suggested dictionary changes are actually incorporated into the official dictionary, the proposed changes must be reviewed by other test raters for validation. Test fairness and reliability depend in part upon the consistency of rater opinion. This is why the scoring dictionaries are so important. As they grow to include the

Table 2: Scoring Dictionary

No.	Scoring Unit Classification	Scoring Unit Type	Scoring Unit	Acceptable Responses	Unacceptable Responses	Notes
5	Names & Numbers	G	July 2nd, 2004	July 2nd, 2004; July 2, 2004; July 2, '04		Must be accurate.
38	Idiomatic Expressions	E	kicked the bucket	Accept any equivalent idiomatic expression in the other language.		If no idiomatic expression exists, accept any accurate interpretation for "dead."
51	Grammar/ Verbs	A	struck her (with his fist)	hit her; punched her	Slapped her	Watch for accuracy in gender and action in context.

various interpretations given by test candidates, raters depend more and more on the dictionaries.

Rating the Consortium Oral Exam

Once the examination has been developed, independently reviewed, and approved by the Consortium's technical committee, it is ready for administration. During the administration process, the oral performance segment of the test that the candidate interprets is tape-recorded. At the conclusion of the examination, the recording, along with other required material, is forwarded to one or more test raters.

Scoring the Exam

The recording of the candidate's performance on the oral segment of the exam makes it possible for raters to rewind the tape when necessary to hear the candidate's rendition of the scoring units. Raters also have a printed version of the transcript of the test text, which they mark while listening to the recording. In terms of marking up the printed script, raters are instructed to:

- Put an "X" through the scoring unit number if the scoring unit is interpreted incorrectly, or is partially or completely omitted.

- Leave the script unmarked if the scoring unit is interpreted correctly.
- Write incorrect or doubtful interpretations near the item for future reference.

Raters also compare their marked scripts with each other. If there is a difference of opinion regarding the marking of a scoring unit, the scoring dictionary is consulted. If the rendition is not found in the dictionary, the raters conduct the necessary research to determine the accuracy of the interpretation. If, after research, the raters still cannot agree, they give credit to the candidate. When that happens, it may be an indication that a scoring unit is not working well and should be replaced or revised. Raters repeat the process through each segment of the test and complete the rater results form with the scores. Scores are then reported to the candidate by the state's program manager.

Passing Scores

Scores are determined based upon the number of scoring units interpreted correctly. The minimum passing score on a Consortium test is 70% for each segment of the test:

- 70% for the two sight translation exercises (minimum of 65% on each segment, with an average of

70% overall).

- 70% on the consecutive test.
- 70% on the simultaneous interpreting test.

The sight translation segments include 25 scoring units each, for a total of 50 scoring units, consisting of approximately 225 words each. Examinees have six minutes to review and translate each document.

The consecutive interpreting test includes 90 scoring units⁶ and consists of 850 to 950 words, 450 to 500 of which are in a language other than English. The consecutive test is timed, and examinees are allowed 22 minutes to interpret witness/attorney question-and-answer testimony. Examinees are allowed to request two repetitions during the consecutive portion of the test.

The simultaneous interpreting test includes 75 scoring units and consists of 800 to 850 words that are pre-recorded in English. The script is recorded at the rate of 120 words per minute, which is very slow, almost artificially so.

Rater Selection and Training

Test raters are selected based upon interpreting experience, testing experience, certifications held, and recommendations from state or federal ➡

court interpreting officials. In many cases, new raters are recommended by existing, well-respected, and knowledgeable raters, program managers, and consultants. Once approved, every new test rater is required to participate in a training session conducted by a qualified test rating supervisor. Here is a brief rundown of what is covered during this training.⁷

Procedures: The test rating supervisor reviews the general test construction theory related to scoring units and how the units are used, along with the mechanics of scoring, the procedures that should be followed for efficient scoring, and the guidelines for completing results report forms. It is critical that the person conducting the training understand the theory of scoring units, how they relate to the scoring unit dictionary, and how the dictionary is to be applied and maintained as part of the test rating process. This is why at least one member of the rater training faculty should be someone who has participated in test construction.

Test content review: Various members of the training group should read the entire test out loud.

Scoring unit and dictionary review: Raters are provided with the scoring dictionary and receive thorough instructions on its use and how to expand its content. The classifications of the scoring units are discussed at length, including what the scoring units are and how they are used within the test scripts, and how to mark the scripts and calculate the final score for each portion of the exam. The test rating supervisor(s) also share tips for efficient and accurate scoring processes and show raters how to complete the results report form. Raters are informed about how much latitude

is associated with the various classifications of scoring units. For example, for dates, names, and times (Type G in Table 2 on page 27), there is no latitude; if the candidate misinterprets the date, name, or time, it is incorrect. This is critical when one considers the

importance of this kind of information during the testimony of a witness. On the other hand, when a test candidate is searching for an equivalent idiomatic expression (Type E in Table 2 on page 27), raters have some leeway when deciding whether the interpretation is

Sites of Interest

Consortium for State Court Interpreter Certification

www.ncsconline.org/d_research/CourtInterp/CICourtConsort.html

Court Interpreters General Information

www.ncsconline.org/D_Research/CourtInterp.html

Federal Court Interpreter Certification Examination

www.ncsconline.org/D_Research/fcice_exam/index.htm

Model Guidelines for Policy and Practice in the State Courts

www.ncsconline.org/wc/publications/Res_CtInte_ModelGuidePub.pdf

National Center for State Courts Publication Page

www.ncsconline.org/D_Research/publications.html



Raters are reminded to respect their colleagues, especially when there is a difference of opinion.

acceptable or not. If the examinee fails to interpret an equivalent idiomatic expression but maintains the meaning of the idiom, the rater is trained to accept that interpretation as correct.

Proper rater practices: An important part of rater training includes a review of proper rater practices. For example, raters are reminded to respect their colleagues, especially when there is a difference of opinion. They are reminded that there is often more than one right way to express a word or phrase. When a test candidate does not use the preferred interpretation, raters are instructed to conduct the necessary research and to acknowledge when the candidate's choice is correct. In addition, the rules regarding regionalisms are reviewed. If an interpretation is widely accepted linguistically in a country or region that speaks the language, raters mark the interpretation as "acceptable." For instance, in various regions of the U.S., Coca-Cola is referred to as "pop," "soft drink," "soda," or "Coke." All of these variants would be considered acceptable (unless a scoring unit is included in the scoring dictionary indicating that the register of the original speaker is to be maintained without variation). The most important point that is emphasized during this training is that raters

must strive for complete neutrality, fairness, and objectivity.

Additional Information

The Consortium's test writing, administration, and rating standards are all available on its website at www.ncsconline.org/d_research/CourtInterp/CICourtConsort.html. Any questions about these standards or requests for more information should be e-mailed to wromberger@ncsc.org.

Notes

1. Since its inception, the Consortium has created and shared with its members a number of other valuable resources to help program managers administer a statewide language services program.
2. Hewitt, William E. *Court Interpretation: Model Guidelines for Policy and Practice in the State Courts* (Williamsburg, Virginia: National Center for State Courts, 1995), www.ncsconline.org/wc/publications/Res_CtInte_ModelGuidePub.pdf.
3. In 2001, standards were created for the development of an "abbreviated" test model, which includes a simultaneous interpreting test (from English into the other lan-

guage) and a spoken English proficiency component.

4. See the test construction manual online at www.ncsconline.org/d_research/CIConsortManuals.html for information about the qualifications of test writers.
5. Specifically, test raters are instructed to "check monolingual and bilingual dictionaries, thesauruses, and websites sponsored by official agencies (not private or individual sites or publications), and official legal publications, whether online or not."
6. Some earlier test forms include 75 scoring units in the consecutive portion of the test. Interpreting 53 of the scoring units correctly would result in a score of 70%.
7. *Test Rating Standards and Resource Materials for Rater Training: Court Interpreting Oral Proficiency Examination* (National Center for State Courts and Consortium for State Court Interpreter Certification, 2001), http://www.ncsconline.org/d_research/CourtInterp/Res_CtInte_12RatersManual.pdf.

ata

Renew your membership online at
www.atanet.org/membership/renew.php

