proficiency, evaluate the speech samples elicited by the COPI using the ACTFL 1999 Proficiency Guidelines—Speaking. The COPI is described as a "low- to medium-stakes test" for formative assessments, that is for measuring student progress (manual, p. 2). Potential users should be advised that the COPI, which was published in 2008, is not fully aligned with current ACTFL proficiency guidelines for speaking because it does not include the “distinguished” level that was added in the 2012 revision (cf. ACTFL, 2012). Further, research evidence that supports the validity and reliability of COPI scores, including interrater reliability, is inadequate. These issues raise the question of the relevance and the usefulness of the COPI as a measure of oral language proficiency in Spanish.

REVIEWER'S REFERENCES


Review of the Computerized Oral Proficiency Instrument (Spanish or Arabic) by ELVIS WAGNER, Assistant Professor, Temple University, Philadelphia, PA, and ANTONY JOHN KUNNAN, Professor, California State University, Los Angeles, CA:

DESCRIPTION. The Computerized Oral Proficiency Instrument (COPI) is an individually administered, semi-adaptive, low- to medium-stakes test of oral proficiency. It is offered in Spanish or Modern Standard Arabic. It was designed to assess adult native-English speakers (including high school students) and provides score reports using the American Council on the Teaching of Foreign Languages (ACTFL) proficiency levels (Novice, Intermediate, Advanced, or Superior). An institution or language program can purchase the COPI from the test developer, the Center for Applied Linguistics (CAL), for $125, which allows for the testing of up to 25 students. Purchasers install the testing program on their computers, and test takers individually take the test on a computer, with no need for an interlocutor, except for an introduction and initial instructions on how to use the computer and the test program.

The test begins with a self-assessment in which test takers choose a can-do statement about their target language oral ability. Based on this self-assessment response, the computer program presents a sample task at the level appropriate for the test taker. The test taker responds to the sample task and then has a chance to change the starting level of the test if he or she thinks the sample task level was too difficult or too easy. Test takers then complete between 7 and 11 tasks based on choices made by both the test taker and the computer algorithm.

The tasks require the test taker to speak in simulated real-life situations. Each task specifies the setting, the situation, and the interlocutor. The lower-level tasks include pictures that provide contextual information. Topics are based on daily life, school, or work contexts. Each task seeks to elicit a certain language function, such as apologizing, and each task is written to elicit language functions at the different ACTFL proficiency levels. The tasks for the Arabic version of the COPI have two versions based on the gender of the test taker; the versions differ with regard to topic and how the test taker is addressed.

For each task, the test taker hears and reads the task directions and then is given from 2 to 4 minutes of thinking time. When the test taker hears the prompt, he or she has 1.5 to 3.5 minutes to speak the response into the computer microphone. Test takers have some control over the difficulty level of the test tasks presented, the choice of topic, and the choice of the language of the directions (in English, or in the target language). The COPI has more than 100 tasks per language, from which the computer algorithm chooses (in conjunction with the preferences of the test taker) the 7 to 11 tasks for each test taker. Once the test is completed, which takes from 45 to 60 minutes, the computer generates a summary showing how many tasks were completed at each difficulty level.

The oral responses are rated by trained human raters. Test takers receive a score following the criteria of the ACTFL Speaking Proficiency Guidelines. The ACTFL Guidelines have 10 possible ratings: Novice Low, Novice Mid, Novice High, Intermediate Low, Intermediate Mid, Intermediate High, Advanced Low, Advanced Mid, Advanced High, and Superior.

DEVELOPMENT. In developing the COPI, the developers report having sought to create a test of oral proficiency based on the ACTFL Speaking
Proficiency Guidelines, which are widely used in academia, government, and language programs, and are recognized as de facto national foreign language proficiency standards. The test developers claim that by using the ACTFL Guidelines, rather than criteria developed specifically for the test, the interpretation of the scores is easier and more transparent for test users.

The COPI is an adaptation of the Simulated Oral Proficiency Interview (SOPI), which is itself an adaptation of the Oral Proficiency Interview (OPI). The OPI was developed to create a performance-based assessment of oral proficiency. The rationale for creating a computer-based version of the OPI was to be able to utilize multimedia computer technology to give test takers more control over the testing situation, while still preserving the standardization of the testing procedure, as well as to increase efficiency and reliability in scoring.

TECHNICAL.

Standardization. Because the COPI scores are based on the ACTFL Guidelines, in which proficiency is defined in terms of the language tasks or functions that a test taker can perform, the test is not a norm-referenced test, and thus it is argued that a standardization/norm sample is not applicable.

Reliability. It is difficult to estimate how reliably the COPI assesses test takers' oral proficiency because of the semi-adaptive nature of the exam. The computer algorithm is designed to choose tasks based in part on the ability level of the test taker as well as the need to assess a variety of language functions, and there are more than 100 possible tasks from which the seven to 11 tasks are selected. In addition, the test taker has some control over the tasks that are selected. Therefore, it is not possible to estimate reliability using a simple parallel forms measure. In addition, because the tasks are designed to measure different language functions, in which a particular test taker will likely have varying ability levels, it is not possible to estimate the internal consistency reliability of the test.

Oral performance tests also present difficulties in the reliability of the scoring. Raters are trained using rater training materials provided with the COPI. The potential raters listen to and rate more than 200 authentic examinee responses. After a potential rater completes the CD training, CAL provides a calibration tape with responses that have to be scored. These scores are submitted to CAL for evaluation, and a rater has to achieve a certain accuracy level to become certified. Nevertheless, because the scoring is done "in house," rather than by a central entity, there is no guarantee that test users will measure or assess the performance of individual raters. The test developers recommend double-rater the responses, but ultimately this decision is up to the users of the COPI.

Validity. Numerous validity studies have been conducted to support inferences made from the COPI. Kenyon and Malabonga (2001) reported a rank order correlation of .95 between ratings on the COPI and the SOPI. Malabonga, Kenyon, and Carpenter (2005), examining the effectiveness of the self-assessment component of the COPI, found a moderate to high correlation with actual performance on the COPI and reported 92% of the test takers chose a task at the level recommended by the self-assessment instrument. Malabonga, Kenyon, and Carpenter (2005) also examined the appropriateness of the length of thinking and speaking time of the tasks.

One of the difficulties in using a semi-direct measure of speaking ability (where the test taker speaks into a computer) is the inherent artificiality of the context in that there is no human interaction. Kenyon and Malabonga (2001) examined COPI test takers' attitudes and found that for the most part, test takers had positive views toward the instrument, especially the fact that they had some control over the choice of tasks, difficulty levels, and thinking and response time.

Fairness. An additional area of investigation that complements reliability and validity is fairness: whether the COPI is fair to test takers based on differences such as gender, age, and presence of disability. Gender is an important variable in the Arabic version of the test because male and female test takers receive different topics for speaking. In this case it would be necessary to know whether the tasks of the two versions are of comparable difficulty. Similarly, because test takers from different age groups take both tests, it would be necessary to know whether the test tasks are of comparable difficulty to the different age groups of test takers. Finally, as more test takers with disabilities are taking such tests, it would be necessary to know what accommodations are provided to test takers with disabilities and whether any studies have been conducted to verify whether the accommodations provided are
not biased in favor of test takers with disabilities. These studies would enhance the beneficial value of the COPI.

COMMENTARY. The COPI is a practical and useful test for the assessment of oral proficiency in Spanish and Arabic. This is very important because there are few other measures of oral proficiency in these languages that are accessible and reasonably priced. The fact that test takers have to speak to a computer, rather than interacting with a human interlocutor, is potentially problematic. Nevertheless, although the COPI is a semi-direct measure, it still is a performance assessment in which test takers actually demonstrate their oral ability. Using human interlocutors for each individual test taker would dramatically increase the cost of the assessment and would also negatively affect the standardization and reliability of the assessment. The COPI seeks to avoid these problems by providing standardized testing procedures while using computer technology to provide important contextual information. Perhaps the key strength of the COPI is the fact that it is semi-adaptive. Using computer technology allows for the individual test taker to have some control over the difficulty level of the tasks and the topics chosen, which should lower negative test-taker affect. The technology also makes it possible to create an algorithm that chooses tasks based on ability level in order to best assess the test taker's proficiency according to the ACTFL guidelines.

Although a number of validity studies have been carried out, more studies (such as fairness) are needed to examine whether the test is fair to all test takers, how the test results are actually used by test users, and how the test results compare with independent measures of oral proficiency.

SUMMARY. The COPI is designed to be an individually administered measure of oral proficiency in Spanish or Arabic. It is designed as a semi-adaptive measure so that test takers would be comfortable taking the test, but also so that it would be best able to provide scores within the context of the ACTFL Proficiency Guidelines. The test meets its goal of assessing oral language proficiency in a practical and useful manner.

REVIEWERS' REFERENCES

Conflict Tactics Scales

Purpose: Designed to obtain reports of domestic violence among abusive adult partners and "physical maltreatment and neglect of children by parents, as well as nonviolent modes of discipline."

Population: Adults.

Publication Date: 2003.

Administration: Individual.

Price Data, 2012: $87.50 per complete test kit including Handbook (146 pages), 10 CTS2 AutoScore forms, and 10 CTSPC AutoScore forms; $45 per 25 CTS2 AutoScore forms; $45 per 25 CTSPC AutoScore forms; $60 per Handbook; $60 per continuing education questionnaire and evaluation form.

Authors: Murray A. Straus, Sherry L. Hamby, and W. Louise Warren.

Publisher: Western Psychological Services.

Acronym: CTS2.


Time: (10–15) minutes.

Acronym: CTSPC.

Scores, 6: Nonviolent Discipline, Psychological Aggression, Physical Assault, Weekly Discipline, Neglect, Sexual Abuse.

Time: (10) minutes.

Review of the Conflict Tactics Scales by M. SHEGHAN DAVIDSON, Assistant Professor of Counseling Psychology, University of Nebraska-Lincoln, Lincoln, NE.

DESCRIPTION. The Conflict Tactics Scales (CTS) are composed of the Revised Conflict Tactics Scales (CTS2) and the Conflict Tactics Scales: Parent-Child Version (CTSPC). These scales are based on the original CTS first published in the early 1970s (Straus, 1973) and developed for epidemiological studies. Both the CTS2 and the CTSPC are self-report inventories, intended to measure the use of specific tactics employed by an intimate couple during a conflict and current maltreatment from parents, respectively. The test authors note that neither the CTS2 nor the CTSPC is intended to measure attitudes, causes, or consequences of conflict or violence. The CTS2 was designed to have symmetry of measurement such that the behavior of both the respondent and the respondent's partner is assessed. The CTSPC was developed as an alternative to retrospective accounts by adults regarding experiences during