that ranged from .86 to .98 for various combinations of raters and forms, indicating that, insofar as the live interviews are valid, the CST probably is too.

To summarize, we found most of the information in the Examinee Handbook and Official Test Manual to be adequate for examinees and proctors to successfully negotiate the test administration (as long as those examinees speak English). However, the Technical Information section is much weaker: (a) the scale and its relationship to scoring are never adequately explained; (b) the small sample sizes and selection procedures limit generalizability; (c) the lack of descriptive statistics makes it difficult to interpret all statistics; (d) the inability to assess the abilities of novice–low, novice–mid, and novice–high students (the levels of most students in U.S. universities) severely limits the CST’s usefulness; (e) claims for four parallel forms overlook the need for equivalent means and variances; and (f) the validity arguments are convincing only under certain fairly prescribed circumstances. Perhaps most problematic, throughout the CST documentation, references are made to other documents (e.g., Clark & Li, 1986) for information that should have been provided in the testing materials themselves.

REVIEWER’S REFERENCE


Review of the Chinese Speaking Test, 1995 Edition by ANTONY JOHN KUNNAN, Assistant Professor of Education, Division of Educational Foundations and Interdisciplinary Studies, California State University, Los Angeles, CA.

The purpose of the Chinese Speaking Test (CST) is “to evaluate the level of oral proficiency in Chinese attained by American and other English–speaking learners of Chinese” (manual, p. 1). The CST is intended for students at proficiency levels from Intermediate to Superior following the guidelines developed by the American Council on the Teaching of Foreign Languages (ACTFL). The uses of the test could include, among others, admission to, placement within, or exemption from a Chinese study program; application for scholarship or appointment; competency testing upon exit from a Chinese program; or certification of Chinese language proficiency for occupational purposes (manual, p. 3).

FEATURES. The CST is a semidirect test, which means the test is administered via an audio–taped recording and a test booklet that contains written materials, maps, and pictures, and the test taker’s responses are recorded on another audio–tape and then rated. It can be administered individually or to a large group in a language laboratory in two forms: The long form lasts 45 minutes and consists of four types of questions designed to elicit speech samples of about 20 minutes duration; the short form lasts about 25 minutes and elicits speech samples of about 15 minutes duration. In Part 1 of the test, the respondent takes part in a personal conversation through several questions about his or her family and education; in Part 2 the test taker is instructed to answer and ask questions, describe a place or activity, give directions, and narrate a story based on a series of pictures; in Part 3 the test taker is instructed to talk about five different topics (two in the short form) and in Part 4 (only in the long form) the test taker is instructed to carry out a specific task based on real life situations.

The recorded speech samples are scored by specially trained raters. The score report contains the test taker’s oral proficiency rating, generally based on the ACTFL proficiency guidelines and scale, ranging from Novice to Superior.

TEST DEVELOPMENT. The test manual contains the procedures adopted for the development of the CST, such as the administration of the preliminary version of the four different forms to two groups of students of Chinese, and the minor modifications made to the overall format and content. The test was revised in 1994 to include more contextualized clues. However, it is unclear how the authors arrived at the four parts to the test; whether they based their decisions on content validity, whether they followed the Oral Proficiency Interview (OPI) format, or whether they had corroborated from any statistical procedures. With regard to level of difficulty, the manual authors state the long form of the test is most appropriate for learners between the levels of Intermediate–High and Superior and the short form for learners between the levels of Novice–High to Intermediate–High on the ACTFL scale. In terms of speededness, responses to questionnaires collected from validation studies indicated that adequate time is allotted for most test takers to respond to each question.

VALIDITY. Three types of validity evidence were gathered: content validity, criterion–related validity, and construct validity. Evidence for content validity was found in the tasks: Test takers are asked to speak in Chinese, they are not required to decode
Chinese, and the tasks closely parallel the OPI format. In a validation study, test takers were asked to indicate on a questionnaire if the questions in each of the tests (the OPI and the CST) adequately probe their maximum level of speaking ability: for the CST, 70% said "yes" and 30% said "no"; for the OPI, 74% said "yes" and 26% said "no," thus providing evidence of the content validity of the CST as a surrogate OPI.

Evidence for criterion-related validity of the 1988 version of the test was offered through correlations obtained from ratings of test performance by two raters who provided independent ratings of the four forms of the CST with the OPI; these correlations ranged from .92 to .98. In a further study, correlations were obtained from ratings by one rater who rated test performance of a test taker on the CST and another rater on the OPI; these slightly lower correlations ranged from .86 to .93. For these two sets of studies, sample sizes were 15 or 16 subjects. No direct correlational study has yet been reported for the revised test.

With regard to construct validation, no separate study of the CST is reported. Instead, the manual authors note that evidence from the content validation and criterion validation studies is sufficient to support the construct validation of the test as a measure of speaking proficiency. In addition, the reader is directed to construct validation studies of the OPI reported in journals.

Despite this excellent case for test validation, a major concern still exists: Why does the test use different tasks? This question should be asked as it is left unclear whether the test developers believe the oral discourse expected is going to be significantly different in each of the tasks. For example, is the oral discourse of giving directions different from giving detailed descriptions or is oral topical discourse different from oral discourse in a specific situation? The manual authors do not provide any discussion on this matter. One empirical method that could be used to establish the basis for the four tasks is to conduct an exploratory factor analysis of test performance data and to interpret the factors (see Kinnan, 1992; Bachman, Davidson, Ryan, & Choi, 1995).

**RELIABILITY.** Interrater reliability information was collected for the four parallel forms: The reliabilities were uniformly high ranging from .89 to .97. Parallel form reliability information was also collected: For Forms A and B the reliabilities were .95 and .99 for two raters and .90 for different raters; and for Forms C and D the reliabilities were .95 and .92 for two raters and .91 for different raters. Both these studies indicate that the test yields consistent results and trained raters are able to use the scoring guidelines consistently. In addition to these classical interrater reliability studies, a generalizability study with facets such as age, gender, native language, a decision-study with varying numbers of raters, and a many-facet Rasch measurement (examining relative severity of different raters and relative difficulty of different tasks) would have enhanced the technical strengths of the test (see Bachman, Lynch, & Mason, 1995; Lumley & McNamara, 1995).

**SUMMARY.** The CST is a reasonably efficient alternative to a direct test. Its validity and reliability are somewhat well established, although the sample sizes are small. Support from additional studies will enhance the confidence with which the test can be used.

**REVIEWER’S REFERENCES**


---

**Christensen Dietary Distress Inventory.**

**Purpose:** Designed to provide an objective assessment of the probability of diet contributing to emotional distress.

**Population:** Ages 18 and above.

**Publication Date:** 1990.

**Acronym:** CDDI

**Scores:** Total score only.

**Administration:** Group.

**Price Data:** Available from publisher.

**Time:** (10–15) minutes.

**Comments:** Self-report inventory.

**Author:** Larry Christensen.

**Publisher:** Pro-Health Publications.

**Cross References:** See T4-460 (1 reference).

**TEST REFERENCES**


Review of the Christensen Dietary Distress Inventory by RICHARD F. FARMER, Assistant Professor of Psychology, Idaho State University, Pocatello, ID:

The Christensen Dietary Distress Inventory (CDDI) is a 34-item self-report questionnaire designed to diagnose persons whose emotional distress is partially or completely due to the intake of specific