Review of the Chinese Proficiency Test 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 Proficiency Test (CPT) is "to evaluate the level of proficiency in (Mandarin) Chinese attained by American and other English-speaking learners of Chinese" (Test Interpretation Manual, p. 1). The test is intended for students at a range of proficiency levels from those who have completed one year of Chinese language instruction at the college or university level up through the maximum level of proficiency normally attained by Chinese majors at the completion of their program. 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 (p. 2).

FEATURES. The CPT contains a total of 150 four-option multiple-choice items divided into two sections, Listening Comprehension and Reading Comprehension. In the Listening Comprehension section, the test taker listens to a number of dialogues between two or more native speakers and longer single-speaker passages in Mandarin Chinese. Following each passage, the test taker is asked one or more questions for which he or she must choose the correct answer from the four-option multiple choices printed in the test booklet in English.

In the Reading Comprehension section, the Structure section assesses the test takers' ability to recognize correct structural patterns in written Chinese in 35 items, and the Reading section assesses comprehension in a variety of printed texts. The 55 items in this section deal only with factual content of the passages and require no background in Chinese literature. In addition, in order to make the test equally fair to test takers whose exposure to written Chinese has been based on either traditional or simplified characters, all material in this section is provided in both character systems.

TEST DEVELOPMENT. The manual does not specifically outline the procedures in the development of the CPT. However, in the section on content validity, the manual notes that in constructing test items "certain content domains were postulated as appropriate to the proficiency of an English-speaking learner of Chinese" (p. 7). Representative items were then constructed for each of the domains representing increasing levels of proficiency in Chinese as a second language.

The CPT is a norm-referenced test and the norming population consists of 479 test takers who took the test between 1984 and 1987 at 60 institutions in the U.S. and abroad. The norming group was identified into three norming groups: beginning, intermediate, and advanced. In the manual, a CPT Percentile Rank Table with the rank data on the three groups of students is presented so that test takers can compare their performance with the performance of the norming population. In the Examiner Handbook this table is called the CPT National Norms Table.

RELIABILITY. Reliabilities and standard error of measurement (SEM) of the CPT based on the 1984-1987 data indicate that the test consistently measures performance. The total test reliability (based on KR-20) and SEM reported were .96 and 5.14, respectively.

VALIDITY. In terms of concurrent validity, the manual reports CPT score patterns obtained by test takers at three self-reported proficiency levels (beginning, intermediate, and advanced) as an external criterion that appears to corroborate and reflect the self-reported proficiency levels.

A content validity discussion of the test is also presented. As mentioned above, test items are constructed based on content domains and levels of proficiency. The levels of proficiency postulated are: Level 1 (Survival), Level 2 (Tourist resident in Chinese society), and Level 3 (Foreign professional in Chinese society). For each of these levels, the target language-use situations and topics are as follows. For Level 1, Listening Comprehension materials include understanding oral language heard in a grocery store, train station, etc., and Reading Comprehension materials include street signs, storefront designations, and other "sight" vocabulary. For Level 2, Listening Comprehension materials include short conversations between two people involving a single conversation turn on topics such as planned trips, family problems, etc., and Reading Comprehension materials include newspaper headlines, printed announcements, and simple newspaper and magazine descriptions. For Level 3, Listening Comprehension materials include oral discourse involving two conversational turns dealing with general academic topics such as current events, health care, etc., and
Chinese Speaking Test, 1995 Edition

Reading Comprehension materials include passages on a wide variety of daily topics such as results of an opinion poll, job descriptions, vacation travel, etc. Questions at Level 3 may deal with authors’ opinions and attitudes, and may require inference. With regard to the Structure section, the manual does not provide a listing of type of items for each content domain and level of proficiency.

Though this attempt to match the test materials and the corresponding target language-use situations the test takers are expecting to encounter is valuable, it is unclear how the levels of proficiency are linked to the content domain. This is a problem because a mere listing of target language-use situations and topics cannot indicate the level of proficiency required to be successful in those situations. But what can resolve this problem is a detailed content analysis of the grammatical, textual, functional, and sociolinguistic features of the test items and the expected response that could reveal the abilities necessary for success on those items at each of the levels (see Bachman, Kunnan, Vanniarajan, & Lynch, 1988).

Another serious problem with regard to the validity of the test is that no construct validation study is reported. A factor-analytic study (Hale, Rock, & Jirele, 1989), or a latent trait model analysis (Embreton, 1985), or a structural modeling study (Muthen, 1988; Kunnan, 1994) would be considered essential to claim that the CPT is a construct-valid test. Additionally, no case is made for the total reliance on the multiple-choice test method, now a major debate in measurement literature (see Bennett & Ward, 1993).

SUMMARY. The CPT is a well-designed and promising test with respectable reliability but its total reliance on the multiple-choice test method and its validity claims are both problematic. Additional studies that address these problems can set aside critical validation concerns that linger at present.

REVIEWER’S REFERENCES


