Multistate Standard-setting Technical Report

PRAXIS™ CHINESE (MANDARIN): WORLD LANGUAGE (5665)

Licensure and Credentialing Research

Educational Testing Service

Princeton, New Jersey

May 2012

Copyright ©2012 by Educational Testing Service. All rights reserved. ETS, the ETS logo and LISTENING. LEARNING. LEADING. are registered trademarks of Educational Testing Service (ETS) in the United States of America and other countries throughout the world.

EXECUTIVE SUMMARY

To support the decision-making process for education agencies with regards to establishing a passing score, or cut score, for the Praxis™ Chinese (Mandarin): World Language (5665) test, research staff from Educational Testing Service (ETS) designed and conducted a two-panel, multistate standard-setting study. The study also collected content-related validity evidence to confirm the importance of the content specifications for entry-level Chinese (Mandarin) teachers.

PARTICIPATING STATES

Panelists from 18 states were recommended by their respective education agency to participate. The education agencies recommended panelists with (a) experience, either as Chinese (Mandarin) teachers or college faculty who prepare Chinese (Mandarin) teachers and (b) familiarity with the knowledge and skills required of beginning Chinese (Mandarin) teachers.

RECOMMENDED PASSING SCORE

The recommended passing score for each panel, as well as the average passing score across the two panels, are provided to help education agencies determine an appropriate operational passing score. For the Praxis Chinese (Mandarin): World Language test, the recommended passing score[1] is 65 (out of a possible 98 raw-score points). The scaled score associated with a raw score of 65 is 164 (on a 100 200 scale).

SUMMARY OF CONTENT SPECIFICATION JUDGMENTS

Panelists judged the extent to which the knowledge and skills reflected by the content specifications were important for entry-level Chinese (Mandarin) teachers. The favorable judgments of the panelists provided evidence that the content covered by the test is important for beginning practice.

To support the decision-making process for education agencies with regards to establishing a passing score, or cut score, for the Praxis™ Chinese (Mandarin): World Language (5665) test, research staff from Educational Testing Service (ETS) designed and conducted a two-panel, multistate standard-setting study. The study also collected content-related validity evidence to confirm the importance of the content specifications for entry-level Chinese (Mandarin) teachers. Panelists were recommended by education agencies to participate. The education agencies recommended panelists with (a) experience, either as Chinese (Mandarin) teachers or college faculty who prepare Chinese (Mandarin) teachers and (b) familiarity with the knowledge and skills required of beginning Chinese (Mandarin) teachers.

[1]*Results from the two panels participating in the study were averaged to produce the recommended passing score.*

The two, non-overlapping panels (a) allow each participating state to be represented and (b) provide a replication of the judgment process to strengthen the technical quality of the recommended passing score. Eighteen states (see Table 1) were represented by 37 panelists across the panels. (See Appendix A for the names and affiliations of the panelists.)

Table 1
Participating States and Number of Panelists (Across Panels)

Arkansas (4 panelists)	Rhode Island (2 panelists)
North Carolina (3 panelists)	Maryland (3 panelists)
Hawaii (l panelist)	Tennessee (l panelist)
North Dakota (1 panelist)	Maine (3 panelists)
Kansas (2 panelists)	Utah (l panelist)
New Hampshire (2 panelists)	Mississippi (4 panelists)
Kentucky (1 panelist)	Vermont (l panelist)
New Jersey (3 panelists)	Montana (l panelist)
Louisiana (2 panelists)	Virginia (2 panelists)

The panels were convened in April and May 2012 in Princeton, New Jersey. For both panels, the same processes and methods were used to train panelists, gather panelists' judgments and to calculate the recommended passing scores.

The following technical report is divided into three sections. The first section describes the content and format of the test. The second section describes the standard-setting processes and methods. The third section presents the results of the standard-setting study.

The passing-score recommendation for the Praxis Chinese (Mandarin): World Language test is provided to each of the represented education agency. In each state, the department of education, the board of education, or a designated educator licensure board is responsible for establishing the final passing score in accordance with applicable regulations. The study provides a recommended passing score, which represents the combined judgments of two groups of experienced educators. The full range of a education agency's needs and expectations cannot likely be represented during the standard-setting study. Each state, therefore, may want to consider the recommended passing score (as well as the separate panels' recommended passing scores) and other sources of information when setting the final Praxis Chinese (Mandarin): World Language passing score (see Geisinger & McCormick, 20 I0). A state may accept the recommended passing score, adjust the score upward to reflect more stringent expectations, or adjust the score downward to reflect more lenient expectations. There is no *correct* decision; the appropriateness of any adjustment may only be evaluated in terms of its meeting the state's needs.

Two sources of information to consider when setting the passing score are the standard error of measurement (SEM) and the standard error of judgment (SEJ). The former addresses the reliability of the Praxis Chinese (Mandarin): World Language test score and the latter, the reliability of panelists' passing-score recommendation. The SEM allows a state to recognize that a Praxis Chinese (Mandarin): World Language test score-any test score on any test-is less than

perfectly reliable. A test score only approximates what a candidate *truly* knows or *truly* can do on the test. The SEM, therefore, addresses the question: How close of an approximation is the test score to the *true* score? The SEJ allows a state to consider the likelihood that the recommended passing score from the current panels would be similar to the passing scores recommended by other panels of experts similar in composition and experience. The smaller the SEJ the more likely that another panel would recommend a passing score consistent with the recommended passing score. The larger the SEJ, the less likely the recommended passing score would be reproduced by another panel.

In addition to measurement error metrics (e.g., SEM, SEJ), each state should consider the likelihood of classification error. That is, when adjusting a passing score, policymakers should consider whether it is more important to minimize a false positive decision or to minimize a false negative decision. A false positive decision occurs when a candidate's test score suggests he should receive a license/certificate, but his actual level of knowledge/skills indicates otherwise (i.e., the candidate does not possess the required knowledge/skills). A false negative occurs when a candidate's test score suggests that she should not receive a license/certificate, but she actually does possess the required knowledge/skills. The state needs to consider which decision error may be more important to minimize.

OVERVIEW OF THE PRAXIS CHINESE (MANDARIN): WORLD LANGUAGE TEST

The Praxis Chinese (Mandarin): World Language *Test at a Glance* document (ETS, in press) describes the purpose and structure of the test. In brief, the test measures whether entry-level Chinese (Mandarin) teachers have the knowledge and skills believed necessary for competent professional practice. The three-hour test is divided into four separately timed sections:

- Section I: Listening with Cultural Knowledge (50 minutes) - 36 multiple-choice questions[2]

- Section II: Reading with Cultural Knowledge (50 minutes) - 39 multiple-choice questions[3]

- Section III: Writing (60 minutes) - Four constructed-response tasks

- Section IV: Speaking (20 minutes) - Four constructed-response tasks

The reporting scale for the Praxis Chinese (Mandarin): World Language test ranges from 100 to 200 scaled-score points. The first national administration of the new test will occur in Fall 2012.

[2] *Six of the 36 multiple-choice questions are pretest questions and do not contribute to a candidate's score.*
[3] *Seven of the 39 multiple-choice questions are pretest questions and do not contribute to a candidate's score.*

PROCESSES AND METHODS

For both expert panels, the same processes and methods were used to train panelists, gather panelists' judgments and to calculate the recommended passing scores. The following section describes the standard-setting processes and methods. (The agenda for the panel meetings is presented in Appendix B.)

The design of the standard-setting study included two non-overlapping expert panels. The training provided to panelists as well as the study materials were consistent across panels with the exception of defining the Just Qualified Candidate (JQC). To assure that both panels were using the same frame of reference when making question-level standard-setting judgments, the JQC definition developed through a consensus process by the first panel was used as the definition for the second panel. The second panel did complete a thorough review of the definition to allow panelists to internalize the definition. The processes for developing the definition (with Panel 1) and reviewing/internalizing the definition (with Panel 2) are described later, and the JQC definition is presented in Appendix C.

The panelists were sent an e-mail explaining the purpose of the standard-setting study and requesting that they review the content specifications for the test (included in the *Test at a Glance* document, which was attached to the e-mail). The purpose of the review was to familiarize the panelists with the general structure and content of the test.

The standard-setting study began with a welcome and introduction by the meeting facilitator. The facilitator explained how the test was developed, provided an overview of standard setting, and presented the agenda for the study.

REVIEWING THE TEST

The first activity was for the panelists to "take the test." (Each panelist had signed a nondisclosure form.) The panelists responded to the multiple-choice questions and constructed-response tasks in a computer-delivered environment similar to environment candidates will be tested under. The purpose of "taking the test" was for the panelists to become familiar with the test format, content, and difficulty.

The panelists then engaged in a discussion of the major content areas being addressed by the test; they were also asked to remark on any content areas that they thought would be particularly challenging for entering Chinese (Mandarin) teachers, and areas that addressed content that would be particularly important for entering Chinese (Mandarin) teachers.

DEFINING THE JUST QUALIFIED CANDIDATE

Following the review of the test, panelists internalized the definition of the Just Qualified Candidate (JQC). The JQC is the test taker who has the minimum level of knowledge and skills believed necessary to be a qualified Chinese (Mandarin) teacher. The JQC definition is the operational definition of the passing score. The goal of the standard-setting process is to identify the test score that aligns with this definition of the JQC.