

Information/Action

Professional Services Committee

Recommended Initial Passing Standards for the California Teacher of English Learners (CTEL) Examination

Executive Summary: This report provides the Commission with recommendations relevant to the determination of passing standards for the California Teacher of English Learners (CTEL) examination.

Staff Direction: Commission staff is seeking direction from the Commission regarding passing standards for the CTEL examination.

Presenters: Amy Jackson, Administrator, Professional Services Division and Dr. Martin B. Karlin, Senior Area Director, National Evaluation Systems, Inc.

Strategic Plan Goal: 1

Promote educational excellence through the preparation and certification of professional educators.

- ◆ Sustain high quality standards for the preparation of professional educators.
- ◆ Sustain high quality standards for the performance of credential candidates.

Recommended Initial Passing Standards for the California Teacher of English Learners (CTEL) Examination

Introduction

This report describes the standard setting studies for the California Teacher of English Learner (CTEL) Examination and provides staff-recommended initial passing standards for each subtest of the examination.

Background

In spring 2004, the Commission's Executive Director appointed an advisory panel of educators with expertise in instruction of English learners to advise Commission staff on the development of a new examination for that pedagogical area. The examination is called the California Teacher of English Learners (CTEL) examination. The CTEL has been developed under the provisions of Assembly Bill 1059 (Ducheny). National Evaluation Systems, Inc. (NES), the Commission's CTEL testing contractor, and Commission staff have worked with this panel to complete this work. The advisory panel, called the English Learner Instruction Design Team (ELIDT) consisted of directors of local Bilingual Teacher Training Programs; a classroom teacher of English learners; a school administrator; and English learner education specialists from school districts, county offices of education, and postsecondary institutions.

In the summer of 2004, the ELIDT developed the draft knowledge, skills, and abilities (KSAs) to be assessed on the CTEL. In the fall of 2004, NES conducted a validity study of the KSAs by distributing surveys to English learner educators throughout the state. The results of the survey were reviewed by the ELIDT and their recommendation for the CTEL KSAs was finalized.

NES also conducted an alignment study to demonstrate that the CTEL KSAs are aligned with the English Language Development Standards for California Public Schools, the Standards of Quality and Effectiveness for Professional Teacher Preparation Programs, the Standards of Quality and Effectiveness for Professional Teacher Induction Programs, and the Teaching Performance Expectations (TPEs). Commission staff reviewed the study conducted by NES and conducted an independent study that confirmed the work conducted by NES.

The CTEL KSAs were adopted by the Commission at its June 1, 2005 meeting.

Using the approved KSAs, multiple-choice and constructed-response items for the CTEL were drafted, reviewed, and revised as needed by both the Bias Review Committee and the ELIDT. Once these items were field-tested, a panel consisting of some members of the ELIDT and some new individuals with appropriate backgrounds in English learner instruction selected marker

responses and scored the constructed-responses from the field test. Additionally, a test guide including the KSAs, test structures, and sample questions was developed to assist candidates in preparing to take the new CTEL.

On December 10, 2005, the first test administration of this new examination was conducted. On January 10-11, 2006 the standard setting study for the CTEL was held in Sacramento to determine the initial passing standard recommendations of California educators.

The CTEL: California Teacher of English Learners Examination

The new CTEL is comprised of subtests differentiated by content area. The subtests consist of both multiple-choice and constructed-response items which are scored using a four-point scale. Constructed-response performance characteristics and scoring scales are provided in Appendix A. The test structure for the CTEL is shown in Table 1 of this agenda item.

The CTEL subtests are administered in two three-hour testing sessions. Subtests I and III are available in the first session and Subtest II is available in the second session. Examinees can choose to take any one or all subtests within a single testing day. Individual subtests are not timed. Individual subtests are not timed. The CTEL will be administered two times each year. For the first administration of the CTEL in December 2005, 527 examinees took Subtest I, 481 examinees took Subtest II, and 600 examinees took Subtest III. Of these, 465 examinees took all three subtests. On January 10-11, 2006 Commission staff and NES conducted a standard setting study for the new examination. The standard-setting procedures used and the results of these studies are described in Part II of this report.

Table 1: Subtest Structure of the CTEL

Subtest	Domains	Number of Multiple-Choice Items	Number of Constructed-Response Items
I	Language Structure and Use	25	none
	First- and Second-Language Development and Their Relationship to Academic Achievement	25	1
	<i>Subtest total</i>	50	1
II	Assessment of English Learners	15	1*
	Foundations of English Language/Literacy Development and Content Instruction	25	1
	Approaches and Methods for ELD and Content Instruction	20	*
	<i>Subtest total</i>	60	2
III	Culture and Cultural Diversity and Their Relationship to Academic Achievement	20	1*
	Culturally Inclusive Instruction	20	*
	<i>Subtest total</i>	40	1
	<i>Total Items</i>	150	4

*One constructed-response item assesses both domains

The Standard Setting Study

The standard setting study for the CTEL was conducted on January 10-11, 2006 with the standard setting panel of English learner specialists. Information about the panel is shown in Appendix B. The purpose of the standard setting procedure is to provide the Commission with recommendations, based on the informed judgments of California educators and subject specialists, relevant to the determination of the initial passing standards. A total of 13 panel members selected from across the state who have varied backgrounds related to English learners, including college and university faculty, and public school educators participated in the study.

Each standard setting study began with an orientation and training session. Panel members were provided the CTEL Knowledge, Skills and Abilities (KSAs) and the subtest forms used for the December 2005 test administration. To help the panel members become familiar with the examinations, the knowledge and skills associated with the items, and the perspective of the examinee, panel members were asked to take the test under simulated test-like conditions. They were asked to read and answer each item independently, without reference to the answer key, and then to score their own performance on the multiple-choice items.

Panel members were then asked to consider the “just acceptable” candidate. Although many of the examinees will exceed the level of knowledge and skills of the acceptably qualified candidate, none should fall below that level. For this reason, panel members were trained to make judgments based on candidates just at the level of knowledge and skills important for effective job performance as a beginning teacher of English learners in a California public school classroom.

After extensive training and the simulated test taking, panel members were asked to complete three rounds of standard setting tasks based on the test structure. This process is briefly described below. A detailed description of the process is found in Appendix C.

In Round One, panel members were asked to individually rate each item on each subtest. They were asked to rate the percent of correct responses that would be expected from a group of “just acceptable” candidates for each multiple-choice item and the level of response that would be achieved by the “just acceptable” candidate for each constructed-response item.

Using the item statistics produced from Round One to inform judgments, Round Two moved the panel from individual item ratings to ratings at the component level (i.e., multiple-choice component and constructed-response component). They were asked the number of multiple-choice items that would be answered correctly and the total score points that would be achieved on the constructed-response items. Panel members were also asked to consider the “component score combination rule”, or the percentage of points that should be allocated to each component (e.g., 80% multiple-choice and 20% constructed-response, 70% multiple-choice and 30% constructed-response).

In the final round of ratings, the panel members were given the results of their Round Two ratings, along with information about the percent of examinees passing at various combinations of scores on the subtest components. They were then asked to make independent recommendations for a passing standard for each component and “component score combination rule”.

Results

Following the standard setting studies, NES calculated for each subtest the median and the distribution of individual Round Three panel recommendations for the multiple-choice and constructed-response test components. Panel recommendations on component score combination rules were also tabulated.

A summary of the panel-based passing score recommendations, including the number of scorable items and the weighting of each component in the total subtest score, is provided in Table 2 below.

Table 2: Panel-Recommended Initial Passing Standards for the CTEL

Subtest	Item Type ^{1/}	Scorable Items	Possible Score Points	Computed Median Based on Panel Recommendations	Component Score Combination Rule ^{2/}	
					80/20	70/30
CTEL I : Language and Language Development	MC	40	40	28.09	2	11
	CR	1	8	5.96		
CTEL II: Assessment and Instruction	MC	48	48	34.11	0	13
	CR	2	16	19.99		
CTEL III: Culture and Inclusion	MC	32	32	24.06	2	11
	CR	1	8	5.43		

^{1/} MC = multiple-choice, CR = constructed-response

^{2/} The component score combination rule is formatted as multiple-choice percent/constructed-response percent (e.g., 80/20 is 80% multiple choice / 20% constructed response).

Staff-Recommended Initial Passing Standards

Based on the guidelines for the establishment of CTEL standards (see Appendix D), staff recommends that the Commission adopt the initial passing standards for the subtests of the CTEL forms administered in December 2005 that:

- are equivalent to the raw score points on the multiple-choice component and on the constructed-response component as shown in Table 3;
- are based on the component score combination rules as shown in Table 3; and
- reflect passing standards that are equivalent for future forms of the test.

For the CTEL, it is appropriate to review passing standards periodically to verify that the standards are fulfilling the responsibility of the Commission to award teaching credentials only to those candidates who have fulfilled requirements related to the teaching of English learners. Following further review, recommendations for any change in the standards will be presented to the Commission for consideration and adoption.

Passing status will be determined on the basis of total subtest performance. Test results will be reported as scaled scores. A scaled score is based on the number of raw score points earned on each component (i.e., multiple-choice and/or constructed-response) and the weighting of each component. For the CTEL, raw scores are converted to a scale from 100 to 300, with a score of 220 representing the passing score as set by the Commission. Scaled scores are used to help ensure

that the level of competence required to pass a given test is independent of the particular form of the test taken.

If the Commission adopts the staff-recommended initial passing standards, as indicated in Table 3, NES will release score reports for the December 10, 2005 test administration by February 22, 2006. The next test administration of CTEL is scheduled for June 24, 2006.

Table 3: Staff-Recommended Initial Passing Standards for the CTEL

Subtest	Multiple-Choice Raw Score Points	Constructed-Response Raw Score Points	Component Score Combination Rule MC/CR	Passing Rate for December 10, 2005 Test Administration by Subtest	Overall Passing Rate for December 10, 2005 Test Administration*
Subtest I	28	6	70/30	74%	54%
Subtest II	34	10	70/30	79%	
Subtest III	24	5	70/30	67%	

* Percent of examinees taking and passing all three subtests at the December 10, 2005 administration

APPENDIX A

PERFORMANCE CHARACTERISTICS AND SCORING SCALE

CTEL SCORING RUBRIC

PERFORMANCE CHARACTERISTICS

PURPOSE	The response fulfills the charge of the assignment in relation to the relevant CTEL Examination Knowledge, Skills, and Abilities
APPLICATION OF CONTENT	The response accurately and effectively applies the relevant CTEL Examination Knowledge, Skills, and Abilities.
SUPPORT	The response is supported with appropriate examples and/or sound reasoning reflecting an understanding of the relevant CTEL Examination Knowledge, Skills, and Abilities

SCORE SCALE

SCORE POINT	SCORE POINT DESCRIPTION
4	<p>The "4" response reflects a thorough command of the relevant CTEL Examination Knowledge, Skills, and Abilities</p> <ul style="list-style-type: none"> • The response completely fulfills the purpose of the assignment by responding fully to the given task. • The response demonstrates an accurate and highly effective application of the relevant knowledge and skills. • The response provides strong support with high-quality, relevant examples and/or sound reasoning.
3	<p>The "3" response reflects a general command of the relevant CTEL Examination Knowledge, Skills, and Abilities.</p> <ul style="list-style-type: none"> • The response generally fulfills the purpose of the assignment by responding to the given task. • The response demonstrates a generally accurate and effective application of the relevant knowledge and skills. • The response provides support with some relevant examples and/or generally sound reasoning.
2	<p>The "2" response reflects a limited command of the relevant CTEL Examination Knowledge, Skills, and Abilities.</p> <ul style="list-style-type: none"> • The response partially fulfills the purpose of the assignment by responding in a limited way to the given task. • The response demonstrates a limited, partially accurate and partially effective application of the relevant knowledge and skills. • The response provides limited support with few examples and/or some flawed reasoning.
1	<p>The "1" response reflects little or no command of the relevant CTEL Examination Knowledge, Skills, and Abilities.</p> <ul style="list-style-type: none"> • The response fails to fulfill the purpose of the assignment. • The response demonstrates a largely inaccurate and/or ineffective application of the relevant knowledge and skills. • The response provides little or no support with few, if any, examples and/or seriously flawed reasoning.
U	The response is unscorable because it is unrelated to the assigned topic or off-task, illegible, or written in a language other than English, or it contains an insufficient amount of original work to score.
B	The written response is blank.

APPENDIX B

STANDARD SETTING PANEL DEMOGRAPHICS

Appendix B

CTEL: California Teacher of English Learners Standard Setting Panel

Total Number	
Appointed	13
Participated	13
Ethnicity	
African American	0
Asian	2
Hispanic	4
White	3
Other/Not Provided	4
Sex	
Female	6
Male	7
Region	
North	5
South	8
Profession	
Public School Educators	7
College/University Educators	6
Years of Experience	
0-6	0
7-10	0
11+	13
Not Provided	0

APPENDIX C

DETAILED STANDARD SETTING PROCESS

Appendix C Standard Setting Rating Tasks

Round One Standard Setting Ratings

In Round One, panel members independently provided item-by-item ratings, first for the multiple-choice items and then for the constructed-response items.

Multiple-Choice Items

For Round One, panel members were provided the following materials:

- the subtest forms used for the December 2005 test administration;
- the accompanying subtest form item statistics; and
- the Round One Rating Form for multiple-choice items

Round One began with a set of five practice, multiple-choice items for each panel member to rate. This set of items represented a range of item difficulties. Panel members were asked to rate each item by responding to the following question.

Imagine a hypothetical group of candidates for CTEL Examination, each of whom is just at the level of knowledge and skills important for effective job performance as a beginning teacher in a California public school classroom.

What percent of this group would answer the item correctly?

0% – 10%	=	1	51% – 60%	=	6
11% – 20%	=	2	61% – 70%	=	7
21% – 30%	=	3	71% – 80%	=	8
31% – 40%	=	4	81% – 90%	=	9
41% – 50%	=	5	91% – 100%	=	10

Panel members were polled as to how they rated each item and as a panel discussed expected performance of the “just acceptable” candidate and the standard setting procedure.

Following the practice set, panel members began the same rating process with the multiple-choice items used on the December 10, 2005 operational test forms. NES analyzed the individual and group results from these item judgments (percentage of “just acceptable” candidates who would answer the item correctly) for use in Round Two of the standard setting process.

Constructed-Response Items

For Round One of the constructed-response item ratings, panel members were provided the following materials:

- the subtest form used for the December 2005 test administration;
- A Constructed-Response Reference Booklet, containing materials for all three subtests, which included:
 - CTEL Performance Characteristics and Scoring Scale, and

- sample examinee responses at each score point for each constructed-response item (“marker responses”)
- the Round One Constructed-Response Item Rating Form.

For Round One of the constructed-response ratings, panel members were asked to rate each item by responding to the following question.

Imagine a hypothetical candidate who is just at the level of knowledge and skills important for effective job performance as a beginning teacher of English Learners in a California public school classroom.

For this constructed-response item, which of the points on the scoring scale represents the level of response that would be achieved by this individual?

Panel members rated the four constructed-response items used on the December, 2005 test forms. In responding to the standard setting question, panel members were asked to refer to the CTET performance characteristics and score point descriptions for the constructed-response items. They were also asked to refer to the marker responses for each score point for each assignment. NES analyzed the individual results from these item judgments for use in Round Two of the standard setting process.

Round Two Standard Setting Ratings

Round Two of the standard setting process moved the panel from providing ratings at the item level to ratings made at the component level (i.e., the multiple-choice component and the constructed-response component) of each subtest. Panel members were asked to provide, for each subtest, (1) separate preliminary passing score recommendations for the set of multiple-choice items and the set of constructed-response items on each subtest and (2) the percent of points to be allocated for each component in the subtest.

For Round Two, panel members were provided the following materials.

- Round One Multiple-Choice Item Rating Summary Sheet, which provided the sum of the item median ratings across all panel members and, for each panel member, the sum of their Round One ratings listed in descending order by score value.
- Round One Constructed-Response Item Rating Summary Sheet, which provided the sum of the item median ratings across all panel members, doubled to reflect the actual combined scores examinees will receive from two scorers. The sheet also provided the sum of each panel member’s Round One constructed-response item ratings doubled to reflect the actual combined scores examinees will receive from two scorers. These individual ratings were listed in descending order by score value.
- Round Two Subtest component Standard Setting Recommendation Form for multiple-choice items.
- Round Two Subtest component Standard Setting Recommendation Form for constructed-response items.

Multiple-Choice Items

Panel members were given an opportunity to discuss the results of the Round One ratings and to provide their thoughts on the merits of various multiple-choice component “cut scores” at the subtest level (understanding that candidates will not “pass” the multiple-choice component alone; candidates’ pass/fail status will be determined at the subtest level, which involves the combination of multiple-choice component and constructed-response component performance). The concept of the multiple-choice component “cut score” was used as a temporary convenience to discuss the aggregated panel member ratings.

Working independently, and considering their own aggregated rating from Round One and the group median, each panel member provided a Round Two multiple-choice component “cut score” recommendation for each subtest by responding to the following question.

Imagine a hypothetical candidate who is just at the level of knowledge and skills important for effective job performance as a beginning teacher of English Learners in a California public school classroom.

What is the number of multiple-choice items on the subtest (out of the total number of scorable items on the subtest) that would be answered correctly by this individual?

Constructed Response Items

Panel members were given an opportunity to discuss the ratings and to provide their thoughts on the merits of various constructed-response component “cut scores” at the subtest level (understanding that candidates will not “pass” the constructed-response component alone; candidates’ pass/fail status will be determined at the subtest level, which involves the combination of multiple-choice component and constructed-response component performance). The concept of the constructed-response component “cut score” was used as a temporary convenience to discuss the aggregated panel member ratings.

Working independently, and considering their own ratings from Round One and the results of the group’s ratings, each panel member provided a Round Two constructed-response component “cut score” recommendation for each subtest by responding to the following question.

Imagine a hypothetical candidate who is just at the level of knowledge and skills important for effective job performance as a beginning teacher of English Learners in a California public school classroom.

What is the total score for the constructed-response items on the subtest (out of the total number of score points for this subtest) that would be obtained by this individual?

Combined Component Scores

Panels were provided the concept of combining subtest component scores in terms of determining the percent of the total points available to be allocated to each component of a subtest. Key issues

that are relevant to this determination were discussed, such as the concept of reliability, the length of each component, and the nature of the information about a candidate's knowledge and skills that is to be provided by each component. The following options that were provided to panels members are intended to yield reliable results and are psychometrically defensible.

Panel members were given two alternatives for allocating points consistent with psychometric standards and the structure of each examination: (a) multiple-choice component 80% and the constructed-response component 20% or (b) the multiple-choice component 70% and the constructed-response component 30%. Panel members were given the opportunity to discuss the options, with advice from Commission staff and NES staff.

Following the discussion, each panel member was asked to independently make a recommendation by responding to the following question.

In combining scores on the multiple-choice component and the constructed-response component to yield a total subtest score, what percent of points should be allocated to each component?

Select one of the following:

_____ *80% multiple-choice component and
20% constructed-response component*

_____ *70% multiple-choice component and
30% constructed-response component*

Following this combined component score rating activity, NES collected and analyzed the panel members' recommendations and informed the panelists of the results.

Round Three Standard Setting Ratings

The goal of Round Three of the standard setting process was to produce a passing standard recommendation for each component of each subtest and a set of panel-recommended rules for combining scores from the multiple-choice and constructed-response components.

For Round Three, panel members were provided the following materials:

- Round Two Multiple Choice Results Summary Sheet, which included the panel's computed median, and each panel member's Round Two multiple choice rating listed in descending order by score value;
- Round Two Constructed-Response Results Summary Sheet, which included the panel's computed median, and each panel member's Round Two constructed-response item rating;
- Summary Statistics Reports for each subtest; and
- A Round Three Subtest Standard Setting Recommendation Form for each subtest.

These materials helped to facilitate a discussion among panel members about their ratings, the nature of the examinee sample, the options for combining component scores, the goal of Round Three, the purpose of the CTEL, and the concept of the just-acceptable candidate.

After a discussion, panel members were asked to independently recommend a passing standard and score combination rule for each CTEL subtest by responding to the following question.

Imagine a hypothetical candidate who is just at the level of knowledge and skills important for effective job performance as a beginning teacher of English Learners in a California public school classroom.

What is the number of multiple-choice items on the subtest (out of the total number of scorable items) that would be answered correctly by this individual?

What is the total score for the constructed-response items on the subtest (out of the total number of score points) that would be obtained by this individual?

In combining scores on the multiple-choice component and the constructed-response component to yield a total subtest score, what percent of points should be allocated to each component?

80% multiple-choice component and 20% constructed-response component

OR

70% multiple-choice component and 30% constructed-response component

As the final step to the standard setting study, each panel member was asked to complete independently a meeting evaluation form regarding the training provided and the task in general.

NES compiled the results of the standards setting panel work for use in the determination of the staff-recommended passing standards presented in this report.

APPENDIX D

CTEL STANDARDS SETTING CONSIDERATIONS

CTEL Standards Setting Considerations

As described in the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999), the standard setting process is a key piece of validity evidence supporting a testing program.

Defining the minimum level of knowledge and skill required for licensure or certification is one of the most important and difficult tasks facing those responsible for credentialing. Verifying the appropriateness of the cut score or scores on the tests is a critical element in validity. The validity of the inference drawn from the test depends on whether the standard for passing makes a valid distinction between adequate and inadequate performance. Often, panels of experts are used to specify the level of performance that should be required. Standards must be high enough to protect the public, as well as the practitioner, but not so high as to be unreasonably limiting. Verifying the appropriateness of the cut score or scores on a test used for licensure or certification is a critical element of the validity of test results (p.157).

In making recommendations to the Commission on passing standards for the CTEL, staff considered the following factors and options that affect the standard setting process in determining the staff-recommended passing standards.

Subtest Scoring Model

The subtest scoring model used with CTEL is a non-compensatory subtest model in which all subtests in a subject area must be passed independently. The CTEL ELIDT panel considered this model when determining the subtest structure of the examination.

Professional Judgments

The recommended passing standards for the CTEL are based upon the professional judgments provided by the members of the Standard Setting Panel. Since these panel recommendations are criterion-referenced—based on expert judgment of the minimum required subject matter knowledge for beginning teachers—examinee performance data provides supplemental, though not necessary, information. Performance data is provided to inform those judgments when there are at least 20 examinees.

Standard Error of Measurement

Standard error of measurement is one way to express test reliability and addresses the imprecision of test data. Measurements are not perfectly reliable. In testing, for example, only one score from a single test administration is available for each examinee. An individual examinee's score may, or may not, be the same as the examinee's hypothetical "true score". However, the standard error allows us to determine a range within which the examinee's true score is likely to lie. Within reasonable limits, the standard error of measurement provides a safeguard against placing undue emphasis on a single numerical score. This is just one index of reliability, and should be applied to the standard setting process in combination with other test-specific characteristics.