Assessment and Accountability Comprehensive Center (AACC)
Evaluation of the Technical Adequacy of Evidence of Assessments of English Language Proficiency: Body of Evidence Summary

Assessment: California English Language Development Test (CELDT)

This body of evidence summary reports the results of the evaluation of technical evidence in support of the CELDT, as analyzed against a validated list of technical adequacy criteria. The table below outlines the types of validity, reliability, and bias and sensitivity evidence associated with various phases of test development in the order they are discussed in the summary. The detailed text following this table explains which quality elements met or exceeded quality expectations. It also provides recommendations for additional types of evidence that would provide support for the technical quality of the assessment. Elements of evidence are divided into Tier 1 and Tier 2. Tier 1 elements ought to be parts of a test's body of evidence, considering both phase and type of development. Tier 2 elements are important, but may include elements that are specific to a particular test. For information regarding the evaluation of the assessment's technical evidence and the technical criteria used, refer to the AACC/WestEd report titled *Evaluation of the Technical Evidence of Assessments for Special Student Populations* at http://www.aacompcenter.org (see Special Populations page).

Type	Phase (Number of possible elements)
Construct validity	Test design and development (10)
Content validity	Test design and development (13)
	Scoring (4)
	Field testing (3)
Consequential validity	Test design and development (1)
	Security (1)
	Reporting (4)
Criterion validity	Test design and development (2)
Reliability	Test design and development (11)
	Scoring (2)
Bias and sensitivity	Test design and development (13)

Body of Evidence

The following documents comprise the body of evidence analyzed for this assessment:

- Technical Report for the California English Language Development Test (CELDT) 2000-2001. CTB/McGraw-Hill. Submitted to the California Department of Education on January 10, 2003.
- Technical Report for the California English Language Development Test (CELDT) 2002-2003 Form B. CTB/McGraw-Hill. Submitted to the California Department of Education on December 10, 2003.
- Technical Report for the California English Language Development Test (CELDT) 2003-2004 Form C. CTB/McGraw-Hill. Submitted to the California Department of Education on January 1, 2005.
- California English Language Development Test (CELDT) Improvements (Provided by CTB/McGraw-Hill). Attachment to California State Board of Education Agenda, April 2003.
- California Department of Education. Media Assistance Packet for School Districts/Schools (2004).
- CELDT Assistance Packet for School Districts and Schools (February 2006).
- CELDT Form E Test Results Interpretation Guide.
- CELDT Reporting 2005–2006 Summary Results Information Guide (February 2006).
- CELDT Form E Test Coordinator's Manual.
- CELDT Scoring Rationales.
- Electronic Tabulation, Interpretation, and Placement (ETIP) Cover Letter and Instructions.
- Scoring Tables and Proficiency Level Descriptors.

Across all types and phases, the technical evidence associated with the CELDT received a rating of meeting or exceeding technical quality expectations in 27 of the 64 evidence/method elements. Further description of specific evidence/method elements of technical adequacy follows.

Construct Validity

Test Design and Development

Of the ten evidence/method elements of construct validity in the test design and development phase, seven received a rating of meeting or exceeding technical quality expectations. The seven elements that met or exceeded expectations were

- *test purpose,*
- *population/classification,*
- *equivalence/comparability,*
- *multi-trait/multi-method/subtest inter-correlation,*
- *accommodation,*
- *fidelity,* and
- *standardization.*

Although three of these elements are Tier 1 elements, evidence of *theoretical foundation/ framework* that meets or exceeds expectations also is desired to determine the degree to which construct validity is evident.

Content Validity

Test Design and Development
Of the 13 evidence/method elements of content validity in the test design and development phase, six received a rating of meeting or exceeding technical quality expectations. The six elements that met or exceeded expectations were

- p-*values/point biserials,*
- *IRT/item fit,*
- *test blueprint,*
- *descriptive statistics,*
- *IRT/test fit,* and
- *linking/equating.*

Although five of these are Tier 1 elements, evidence of *alignment (items-to-standards), expert judgment,* and *alignment (test form-to-blueprint)* that meets or exceeds expectations is also desired to determine the degree to which content validity is evident.

Of these elements, p-*values/point biserials* and *descriptive statistics* included information derived from field testing.

Scoring
Of the four evidence/method elements of content validity in the scoring phase, three received a rating of meeting or exceeding technical quality expectations. The three elements that met or exceeded expectations were

- *scale,*
- *standard setting,* and
- *training of scorers/scoring protocol.*

Scale and *standard setting* are the two Tier 1 elements for content validity in the scoring phase.

Of these elements, *scale* and *standard setting* included information derived from field testing.

Field Testing
Of the three evidence/method elements of content validity in the field testing phase, one received a rating of meeting or exceeding technical quality expectations. The element that met or exceeded expectations was

- *blueprint.*

Evidence of one Tier 1 element, *sampling* or *norming* that meets or exceeds expectations is also desired to determine the degree to which content validity is evident.

Consequential Validity

Test Design and Development
The one evidence/method element of consequential validity in the test design and development phase received a rating of meeting or exceeding technical quality expectations. The element that met or exceeded expectations was

- *use of results.*

Reporting
Of the four evidence/method elements of consequential validity in the reporting phase, three received a rating of meeting or exceeding technical quality expectations. The elements that met or exceeded expectations were

- *reporting category,*
- *N,* and
- *central tendency/variation.*

Security
The one evidence/method element of consequential validity in the security phase received a rating of meeting or exceeding technical quality expectations. The element that met or exceeded expectations was

- *protocols.*

Criterion Validity

Test Design and Development
Of the two evidence/method elements of criterion validity in the test design and development phase, c*ross tabulations* and *Pearson correlation,* neither received a rating of meeting or exceeding technical quality expectations.

Evidence of one Tier 1 element *cross tabulations* or *Pearson correlation* that meets or exceeds expectations is desired to determine the degree to which criterion validity is evident.

Reliability

Test Design and Development
Of the 11 evidence/method elements of reliability (stability and consistency, internal consistency, generalizability, and classification consistency) in the test design and development phase, three received a rating of meeting or exceeding technical quality expectations. The elements that met or exceeded expectations were

- *standard error of measurement/confidence intervals,*
- *coefficient alpha,* and
- *classification error.*

Two of these are Tier 1 elements. Evidence of *test-retest* or *alternate form* reliability that meets or exceeds expectations is also desired to determine the degree to which reliability is established.

Of these elements, s*tandard error of measurement/confidence intervals* included information derived from field testing.

Scoring
Of the two evidence/method elements of reliability (inter-rater) in the scoring phase, two received a rating of meeting or exceeding technical quality expectations. The elements that met or exceeded expectations were

- *correlation (kappa)* and
- *percent correspondence.*

Bias and Sensitivity

Test Design and Development
Of the 13 evidence/method elements of *expert review* and *DIF analysis* across seven types of bias and sensitivity (linguistic, ethnicity/race, cultural/religious, geographic, SES, disability, and gender), none received a rating of meeting or exceeding technical quality expectations.

Evidence of *expert review* and analyses of other types of bias are desired to determine freedom from bias.

Preliminary AACC Comments:

Overall, the technical evidence of validity for CELDT met or exceeded expectations, while the evidence of reliability and freedom from bias and sensitivity did not meet expectations in some key areas.

All Tier 1 elements at least met expectations for consequential validity; most Tier 1 elements at least met expectations for two of the remaining three types of validity. For criterion validity, no Tier 1 elements met expectations. For construct validity, one element of test design and development, *theoretical foundation/framework*, did not meet expectations. Tier 1 elements met or exceeded expectations in one phase of content validity, scoring. In test design and development, five elements at least met expectations, but three did not. The evidence for two elements, *IRT/Item Fit* and *Blueprint*, was noted to be particularly comprehensive. For the element *alignment (items-to-standards)*, the documents did not include sufficient information about the degree to which assessment items align with California's English language development standards, and for expert judgment, the documents did not clarify the role that experts played during field test development. No evidence was provided for the element *alignment (test form-to-blueprint).* In the phase field testing, no elements met expectations. For both types of evidence, *sampling* and *norming*, the documents did not provide details about the source (e.g., home language survey-CDE or Common Core of Data) or age (i.e., publishing date) of data used. In addition, it was unclear the degree to which the data apply to the targeted student population and whether language proportions are representative of the targeted student populations.

For reliability, both elements for scoring at least met expectations, and so did two Tier 1 elements for test design and development. For the elements *test-retest* and *alternate form*, analysts noted that although administering the annual and initial forms to the same students may be, technically, a kind of test/retest or alternate forms reliability, this is not the way in which this information was presented in the documents. In addition, while an assertion about test-retest reliability appears in the documents, no method are explicitly described. For bias and sensitivity, none of the elements at least met expectations. For the *expert review* elements *ethnicity/race* and *gender*, the documents mention review for "ethnic, racial" and gender bias/sensitivity, describe the contractor's internal review protocol, and list the review committee panelists. However, neither the protocol for these meetings nor information about the ways in which outcomes from those sessions affect item selection is presented. For the *DIF analyses* elements, only *gender* is specifically mentioned ("Favor Female" or "Against Female"); while the documents make reference to flagging items, no supporting details are provided.

Test Publisher Comments:

Publisher provided no response/additional comment.

Final AACC Comments and Recommendations:

The AACC sent CTB-McGraw-Hill a notification of the pending evidence review and request for additional information. The publisher has not responded to this request for information. These analyses are based upon the information listed at the beginning of this document, available from the Web sites of CTB-McGraw-Hill and the California Department of Education.

The contents of this evaluation summary were developed under a grant from the Department of Education. However, those contents do not necessarily represent the policy of the Department of Education, and you should not assume endorsement by the Federal Government.