



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

Assessment: IDEA Proficiency Test-Oral (IPT-O) 2004

This body of evidence summary reports the results of the evaluation of technical evidence in support of the IPT-O, 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 this 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 a complete list of the Tier 1 and Tier 2 elements, see document: *Assessments for English Language Learners: Technical Adequacy Criteria—Tiers*. For operational definitions of the criteria, see document: *ELL Assessment Technical Criteria—Operational Definitions*. 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 Manual, IPT I Oral, Grades K-6, English, Forms E & F. (2004) Ballard & Tighe.
- Technical Manual, IPT II Oral, Grades 7-12, English, Forms C & D. (2004) Ballard & Tighe.
- Examiner's Manual, IPT I Oral, Grades K-6, English, Forms E & F. (2004) Ballard & Tighe.
- IPT NCLB Compliance Handbook. (2004) Ballard & Tighe. Includes IPT I Oral Examiner's and Technical Manual Addendum & IPT II Oral Examiner's and Technical Manual Addendum.
- Examiner's Manual, IPT I Oral, Grades K-6, English, Forms E & F, Second Edition. (2006) Ballard & Tighe.
- Technical Manual, IPT I Oral, Grades K-6, English, Forms E & F, Second Edition. (2006) Ballard & Tighe.
- Examiner's Manual, IPT II Oral, Grades 7-12, English, Forms C & D, Second Edition. (2006) Ballard & Tighe.
- Technical Manual, IPT II Oral, Grades 7-12, English, Forms C & D, Second Edition. (2006) Ballard & Tighe.
- Examiner's Manual, Pre-IPT Oral, Ages 3-5, English, Third Edition. (2006) Ballard & Tighe.
- Technical Manual, Pre-IPT Oral, Ages 3-5, English, Third Edition. (2006) Ballard & Tighe.

Across all types and phases, the technical evidence associated with the IPT-O (2004) received a rating of meeting or exceeding technical quality expectations in 22 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, five received a rating of meeting or exceeding technical quality expectations. The five elements that met or exceeded expectations were

- *test purpose*,
- *population/classification*,
- *theoretical foundation/framework*,
- *fidelity*, and
- *standardization*.

These include the four Tier 1 elements of *test purpose*, *population/classification*, *theoretical foundation/framework*, and *fidelity*.

## Content Validity

### Test Design and Development

Of the 13 evidence/method elements of content validity in the test design and development phase, one received a rating of meeting or exceeding technical quality expectations. The element that met or exceeded expectations was

- *test blueprint*.

This is a Tier 1 element, but evidence of *alignment (items-to-standards)*, *expert judgment*, *IRT/item fit*, *alignment (test form-to-blueprint)*, *IRT/test fit*, and *linking/equating* that meets or exceeds expectations is desired to determine content validity.

### 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 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 at the scoring phase.

### 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

- *norming*.

*Norming* is one of two Tier 1 elements for content validity at the field testing phase.

## 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*.

*N* and *central tendency/variation* are the two Tier 1 elements for consequential validity at the reporting phase.

## Security

The one evidence/method element of consequential validity in the security phase, *protocols*, did not receive a rating of meeting or exceeding technical quality expectations.

## Criterion Validity

### Test Design and Development

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

- *cross tabulations* and
- *Pearson correlation*.

## 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, four received a rating of meeting or exceeding technical quality expectations. The elements that met or exceeded expectations were

- *standard error of measurement/confidence intervals*,
- *test-retest*,
- *alternate form*, and
- *coefficient alpha*.

*Standard error of measurement/confidence intervals*, *test-retest*, *alternate form*, and *coefficient alpha* are Tier 1 elements of reliability at the test design and development phase.

## 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.

## Preliminary AACC Comments:

In response to notification of the IPT 2004 evidence evaluation and request for additional documentation, Ballard & Tighe responded that the IPT 2004 series of tests are “currently out of date and print.” Additionally, the tests are “specifically used for placement only at the district

and school levels and are not relevant for NCLB Title I and Title III purposes.” The publisher requested that the AACC instead review the evidence associated with their current, NCLB-compliant test materials, and provided the AACC with materials for review. Analysts will review the new IPT materials. However, given that the IPT 2004 is currently approved for use in at least two states<sup>1</sup> and is in use at the local level for placement, a review of the evidence associated with the IPT 2004 was conducted. The evidence presented in this summary represents that review.

### Test Publisher Comments:

The publisher provided a detailed response to the AACC evidence summary (below), including sources of evidence that the publisher believed addressed the technical criteria.

IPT 2004 Oral Tests  
Additional Evidence:

### Content Validity—Test Design and Development

- *alignment (items-to-standards)*: This is and has been available on our web site at [http://www.ballard-tighe.com/assess/alignment\\_assess.html](http://www.ballard-tighe.com/assess/alignment_assess.html)

### Final AACC Comments and Recommendations:

AACC analysts considered the additional evidence provided by the publisher (noted above) vis-à-vis the technical review criteria. When the additional evidence was found to meet or exceed expectations, ratings were revised. This summary reflects the analysts’ final ratings.

This summary is intended to inform consumers and test publishers of the breadth and depth of evidence relevant to supporting an assessment’s validity, reliability, and freedom from bias.

We appreciate the thoughtful comments offered by Ballard-Tighe in reference to the summary evaluation of the IPT-Oral 2004 materials. However, regarding the link to the alignment documentation, no explicit reference to such studies was found in the Technical Manuals. Additionally, no detailed description of the nature of the studies (e.g., independent versus internal; qualifications of judges) associated with these alignment findings was provided. The criteria for this type of evidence are as follows:

#### Alignment

In-process alignment and/or ex post facto alignment studies done (independent); appropriate unit(s) of analysis and model/appropriate dimensions evaluated; explanation of process or results (including limitations). In-process alignment may be done by writers, editors, or other developers and expert reviewers during the item development process. Ex post facto alignment should be done by independent experts in assessment, standards, and relevant content areas. Alignment procedures and studies should look for appropriateness of item content and cognitive level as described in individual standards, and coverage (breadth and depth) as reflected by the set of standards.

---

<sup>1</sup> United States Government Accountability Office. (2006). *No Child Left Behind Act: Assistance from Education Could Help States Better Measure Progress of Students with Limited English Proficiency* (GAO-06-815). Washington, DC: U.S. Government Printing Office.

Finally, specific information about item-to-standard alignments was provided only in relation to the TESOL standards and the California, Colorado, and Virginia proficiency standards. Alignment information was very general (text comments, checkmarks, bullet points) for all other cited studies (Council of Great City Schools/National Clearinghouse Assessment Standards and the proficiency standards for Arkansas, Hawaii, Idaho, Maryland, Missouri, and Oklahoma).

The rating for this additional evidence changed from a “0” to a “1” and is “below expectations.”

---

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.