

**Idaho English Language Assessment  
(IELA)**

**Technical Report**

**2007**



## Table of Contents

	Page
1. Purpose of the Technical Report	5
2. Description of the IELA	5
2.1 Purpose of the IELA	5
2.2 Past and Present IELA Forms	5
2.3 Structure of the IELA	6
3. IELA 2007 Administration	8
3.1 Test Window	8
3.2 Assessment Training	8
3.3 Examiner Scripts	8
3.4 Listening Test Administration	8
3.5 Setting for the Test	8
3.6 Timing	9
3.7 Prompting or Repeating Test Information	9
3.8 Testing Absentees	10
3.9 Testing Accommodations	10
4. Test Security	12
4.1 Test Security Agreement	12
4.2 Bar –coding and Return of Secure Materials	12
4.3 Storage and Shredding of Secure Materials	12
5. IELA 2007 Item Analyses	13
6. Scaling and Equating	16
7. Reliability of the IELA	22
8. Validity of the IELA	27
8.1 Content and Construct Validity	27
8.2 Criterion-related Validity	28

9. Performance by Year	31
10. Item Development	38
10.1 Alignment Study	38
10.2 Item Development	38
10.3 Content and Bias Review	38
Appendix A. Mountain West Assessment Consortium, Introduction to Foundation Document.	39
Appendix B. Plan for IELA Item Development/Proposed Test Blueprints	44
Appendix C. Item Writing Overview and Guidelines, Multiple Choice (MC) and Constructed Response (CR)	78

## **Idaho English Language Proficiency Assessment 2007 Technical Report**

### **1. Purpose of the Technical Report**

The purpose of this report is to provide the Idaho OSBE, educators, citizens, researchers, and other interested parties with technical documentation for the development, administration, and reporting of the 2007 Administration of the IELA. This report includes evidence of the reliability and validity of the assessment as well as information on the appropriate use and interpretation of test scores. Although this technical report covers the 2007 administration of the IELA, some data from a previous administration are included for reference and comparison.

### **2. Description of the IELA**

**2.1 Purpose of the IELA.** The Idaho English Language Assessment (IELA) is an assessment of English language proficiency for grades K-12. It is a modified version of an assessment developed for the Mountain West Consortium and designed to fulfill the requirements of ‘No Child Left Behind’ (NCLB) legislation. The IELA assesses English proficiency in Listening, Speaking, Reading, and Writing and reports scores in each of those language domains as well as in Comprehension (a combination of select items from the Listening and Reading test) and a total score, representing overall English proficiency. The IELA was designed to assess the status of a student’s proficiency in English and to measure progress in attaining English proficiency.

The IELA was designed to be administered to all students identified as ‘limited English proficient’ (LEP) in the State of Idaho. Per the instructions printed in the IELA Examiner Manuals, “An LEP student is an English language learner specifically identified for a language development program and for whom LEP funding was received. Not all English language learners are LEP students; for example, a student may not have been placed in an LEP program, or may have already exited a program.” Districts and schools were also given the option of administering the IELA to their LEPX students who were still within the 2-year monitoring period after exit from an LEP program.

**2.2 Past and Present IELA Forms.** The first set of IELA forms, designated IELA 2006, was developed and administered in spring 2006. These forms were based on Mountain West Form I. More detailed information about these forms is included in the IELA Technical Report, 2006.

A second set of IELA forms, designated IELA 2007, was developed and administered in spring 2007. IELA 2007 forms were similar in structure to the IELA 2006 forms but with approximately 70% different items. The new items on IELA 2007 were drawn from the

Mountain West Consortium item bank (i.e., Forms II and III). New items were reviewed for content and structure and edited where appropriate. Directions for administration were revised, where necessary and appropriate, to conform to the conventions adopted in IELA 2006. Items that were in common between the 2006 and 2007 forms served as anchor items to equate the 2007 to the 2006 forms.

**2.3 Structure of the IELA.** The Idaho English Language Assessment (IELA) is an assessment of English language proficiency for grades K-12. The IELA assesses English proficiency in Listening, Speaking, Reading, and Writing and reports scores in each of those language domains as well as in Comprehension (a combination of select items from the Listening and Reading test) and a total score, representing overall English proficiency. IELA test forms were designed for specific grade/grade clusters, K, 1-2, 3-5, 6-8, and 9-12, as shown in Table 1. For every grade cluster except Kindergarten, there are two forms differentiated by a number suffix (e.g., C1 and C2). The level 1 forms were designed to be administered to students on the lower end of the English proficiency scale (i.e., Beginner) and the level 2 forms designed for students on the upper end of the scale (i.e., Intermediate and Advanced). Within each grade cluster, the Listening and Speaking tests on level 1 and 2 forms are identical (i.e., feature the same items). The Reading and Writing tests on level 1 and 2 forms within a grade cluster are different, both in terms of the numbers of items and the content.

Table 1 shows for each test form, the grade cluster in which it is administered and the numbers of items by item type in each language domain as well as the number of points represented by those items. The items and points in the Comprehension column do not contribute to the Totals shown in the last two columns because all Comprehension items are part of the Listening or Reading tests.

**Table 1. Structure and Content of IELA 2007 Test Forms**

Form	Grade Cluster	Item Type	Listen		Speak		Read		Write		Comp		Total	
			Itm	Pts	Itm	Pts	Itm	Pts	Itm	Pts	Itm	Pts	Itm	Pts
<b>A</b>	<b>K</b>	<b>MC</b>	9	9	-	-	23	23	-	-	16	16	32	32
		<b>SA</b>	13	13	10	10	13	13	-	-	13	13	36	36
		<b>ER</b>	-	-	4	12	-	-	-	-	-	-	4	12
		<b>Total</b>	<b>22</b>	<b>22</b>	<b>14</b>	<b>22</b>	<b>36</b>	<b>36</b>	<b>22*</b>	<b>22*</b>	<b>29</b>	<b>29</b>	<b>94</b>	<b>102</b>
<b>B1</b>	<b>1-2</b>	<b>MC</b>	22	22	-	-	15	15	-	-	31	31	37	37
		<b>SA</b>	-	-	10	10	-	-	11	11	-	-	21	21
		<b>ER</b>	-	-	4	12	-	-	2	4	-	-	6	16
		<b>Total</b>	<b>22</b>	<b>22</b>	<b>14</b>	<b>22</b>	<b>15</b>	<b>15</b>	<b>13</b>	<b>15</b>	<b>31</b>	<b>31</b>	<b>64</b>	<b>74</b>
<b>B2</b>		<b>MC</b>	22	22	-	-	20	20	-	-	39	39	42	42
		<b>SA</b>	-	-	10	10	-	-	10	10	-	-	20	20
		<b>ER</b>	-	-	4	12	-	-	3	10	-	-	7	22
		<b>Total</b>	<b>22</b>	<b>22</b>	<b>14</b>	<b>22</b>	<b>20</b>	<b>20</b>	<b>13</b>	<b>20</b>	<b>39</b>	<b>39</b>	<b>69</b>	<b>84</b>
<b>C1</b>	<b>3-5</b>	<b>MC</b>	22	22	-	-	15	15	4	4	31	31	41	41
		<b>SA</b>	-	-	10	10	-	-	5	5	-	-	15	15
		<b>ER</b>	-	-	4	12	-	-	2	6	-	-	6	18
		<b>Total</b>	<b>22</b>	<b>22</b>	<b>14</b>	<b>22</b>	<b>15</b>	<b>15</b>	<b>11</b>	<b>15</b>	<b>31</b>	<b>31</b>	<b>62</b>	<b>74</b>
<b>C2</b>		<b>MC</b>	22	22	-	-	18	18	9	9	37	37	49	49
		<b>SA</b>	-	-	10	10	1	2	-	-	1	2	11	12
		<b>ER</b>	-	-	4	12	-	-	3	10	-	-	7	22
		<b>Total</b>	<b>22</b>	<b>22</b>	<b>14</b>	<b>22</b>	<b>19</b>	<b>20</b>	<b>12</b>	<b>19</b>	<b>38</b>	<b>39</b>	<b>67</b>	<b>83</b>
<b>D1</b>	<b>6-8</b>	<b>MC</b>	22	22	-	-	15	15	5	5	32	32	42	42
		<b>SA</b>	-	-	10	10	-	-	4	4	-	-	14	14
		<b>ER</b>	-	-	4	12	-	-	2	6	-	-	6	18
		<b>Total</b>	<b>22</b>	<b>22</b>	<b>14</b>	<b>22</b>	<b>15</b>	<b>15</b>	<b>11</b>	<b>15</b>	<b>32</b>	<b>32</b>	<b>62</b>	<b>74</b>
<b>D2</b>		<b>MC</b>	22	22	-	-	18	18	10	10	38	38	50	50
		<b>SA</b>	-	-	10	10	-	-	-	-	-	-	10	10
		<b>ER</b>	-	-	4	12	2	6	3	10	2	6	9	28
		<b>Total</b>	<b>22</b>	<b>22</b>	<b>14</b>	<b>22</b>	<b>20</b>	<b>24</b>	<b>13</b>	<b>20</b>	<b>40</b>	<b>44</b>	<b>69</b>	<b>88</b>
<b>E1</b>	<b>9-12</b>	<b>MC</b>	22	22	-	-	15	15	7	7	32	32	44	44
		<b>SA</b>	-	-	10	10	-	-	2	2	-	-	12	12
		<b>ER</b>	-	-	4	12	-	-	2	6	-	-	6	18
		<b>Total</b>	<b>22</b>	<b>22</b>	<b>14</b>	<b>22</b>	<b>15</b>	<b>15</b>	<b>11</b>	<b>15</b>	<b>32</b>	<b>32</b>	<b>62</b>	<b>74</b>
<b>E2</b>		<b>MC</b>	22	22	-	-	19	19	10	10	39	39	51	51
		<b>SA</b>	-	-	10	10	-	-	-	-	-	-	10	10
		<b>ER</b>	-	-	4	12	2	6	3	10	2	6	9	28
		<b>Total</b>	<b>22</b>	<b>22</b>	<b>14</b>	<b>22</b>	<b>21</b>	<b>25</b>	<b>13</b>	<b>20</b>	<b>41</b>	<b>45</b>	<b>70</b>	<b>89</b>

\* Items on the Kindergarten Writing test are configured as a checklist completed by the examiner.  
 MC - Multiple Choice; SA - Short Answer; ER - Extended Response

### **3. IELA 2007 Administration**

**3.1 Test Window.** The test window for IELA 2007 was February 21 through April 6, 2007. An additional week was granted upon request by school districts that needed to accommodate migrant students.

**3.2 Assessment Training.** To prepare districts for the administration of the spring 2007 IELA, three PowerPoint Presentations were created – *What’s New, Test Administration,* and *Post-Test Instructions.* These documents were posted with complete notes at the Board of Education website ([www.boardofed.idaho.gov/lep/LEPAssessment.asp](http://www.boardofed.idaho.gov/lep/LEPAssessment.asp)). An audio version was also available online.

Each District Test Coordinator was encouraged to read through these presentations prior to administration and to consider using the PowerPoint presentations to train test administrators. In addition, a series of 5 hosted WebEx seminars (November 2 & 29, 2006, and December 1, 4, and 13, 2006) based on those presentations was offered to any and all test coordinators and examiners.

To prepare for testing, examiners were instructed (in the examiner manual) to:

- read the manual completely;
- ensure that they had adequate materials for all students who would be tested;
- notify students in advance of testing;
- affix student barcode labels to answer documents; and
- secure a CD player (or computer with CD-ROM drive, sound card and speakers) for administering the Listening test, and check the CD and the sound quality.

**3.3 Examiner Scripts.** Specific step-by-step instructions and script were provided for each test form in an examiner manual specific to that particular form. Scoring guides were provided for all oral constructed responses. Such items occurred throughout the Kindergarten forms, but only in the Speaking test at all other grade spans. Where appropriate, examples of full-credit and partial-credit responses were provided.

**3.4 Listening Test Administration.** The Listening test was administered from a CD recording. This ensured that all students heard the questions in the same voice and at the same pace. The recording included a chime after each question signaling the examiner to pause the CD while students responded. A printed Listening Script for each form was available to any school that requested it.

**3.5 Setting for the Test.** For the individually administered subtests, examiners were advised as follows: “The test setting should be a quiet one-to-one environment. The testing should take place where other students cannot hear or see the testing materials. The examiner should

sit close enough to the student to point to questions and illustrations in the student's test booklet during test administration."

For the group-administered subtests, examiners were advised as follows: "The test setting for the group-administered sections is a quiet classroom. The students should have in front of them only their test booklet, answer document, and a No. 2 pencil."

**3.6 Timing.** The IELA is an untimed test and examiners were advised to allow students as much time as they needed to finish any given subtest.

**3.7 Prompting or Repeating Test Information.** The following rules regarding prompting or repeating information were printed in all examiner manuals:

*Prompting* is the provision of additional information to students during administration of the assessment. Prompting includes

- elaborating on questions,
- clarifying information provided in reading selections or any test question,
- pointing out specific information in the questions or graphics,
- providing cues that might normally be part of an instructional strategy, and/or
- suggesting strategies that a student may use to arrive at a correct response.

In general, prompting is **not** allowed in this test because it may give an unfair advantage to some students. However, in specific situations where partial or unclear responses are given, the following general prompts are appropriate.

To clarify the student's response, the examiner may say,

*I don't understand what you said.*

*Can you tell me more?*

If the student answers in another language, the examiner may say,

*Can you say that in English?*

The examiner may repeat directions, if necessary, but must do so before the child begins a response.

If there is a distraction or interruption, the selection or question may be repeated.

If a student asks for a question to be repeated, the examiner may repeat the question only once.

If the student still does not understand what is being asked, the examiner should score that question as though the student gave no response (*BL*).

The examiner must not modify directions in any way. To do so would provide an unfair advantage to one student or a group of students over others.

The examiner should allow approximately 15 seconds of wait time for a student to begin a response to a question. This gives the student time to gather his or her thoughts and to think carefully before responding in English. If a student has not responded after 15 seconds, the examiner should move on to the next item or task and score the item as “no response” (*BL*).

**3.8 Testing Absentees.** Examiners were advised to make every effort to see that all LEP students in the school were administered all sections of the IELA. If a student was absent for a particular testing session, a make-up test was to be scheduled, as long as it was within the testing window.

**3.9 Testing Accommodations.** For visually impaired students, the IELA 2008 was available (by special order) in Braille and in Enlarged Print. One (1) Braille forms and six (6) Enlarged Print forms were ordered before the December 5, 2007, deadline.

For deaf and hard-of-hearing students, the following guidelines were printed in all examiner manuals:

Lip-reading for those students who possess this ability may be possible for those parts of the test where the teacher reads the test questions aloud. A copy of the Listening Test Script is available and may be ordered from the IELA Coordinator at [iel@QuestarAI.com](mailto:iel@QuestarAI.com), so that an examiner may administer the Listening Test to a deaf student with lip-reading ability. For the Speaking Test, a deaf student with lip-reading ability must also have the ability to answer in spoken English; otherwise the test should not be administered to him or her. IEP teams should make such determinations on a case by case basis. The Listening and Speaking prompts should not be translated into sign language. Doing so is equivalent to translating into another spoken language, such as Spanish, or Arabic, and thus would invalidate the test. However, those Reading and Writing prompts meant to be spoken by the teacher may be translated into sign language if necessary.

For students with an Educational Learning Plan (ELP) or Individual Education Plan (IEP) on file, the following list of allowable accommodations was printed in all examiner manuals:

- 504 type accommodations (physical disabilities, mobility issues, etc.)
- Separate testing setting, small group or individual administration
- Objects or markers to assist with maintaining place on the page

- Administration of the test at home, in a hospital, or any other required setting by school personnel
- Any additional “non-linguistic” accommodation required that would not interfere with test validity
- Teacher uses highlighters or any similar device to distinguish words or key phrases within text
- Noise buffers
- Breaks within sections, except as these are part of the standard administration procedures (breaks between sections are not controlled)
- Student reads questions aloud to self (must be taking the test in a separate room)
- Repeating questions
- Orally read test questions in English (other than reading passages) or audiotape test questions in cases where student would normally read the question
- Read, reread, paraphrase or simplify test directions in English (not test items or test item directions)
- Explanation of test directions in English (not test items or test item directions)
- Direct translations of test directions into Native Language (not test items or test item directions)
- Sign test directions to students (not test items or test item directions which students would normally read themselves)

However, examiners were warned that such accommodations should be used only when absolutely necessary and only with students with an ELP or IEP on file.

Certain accommodations would necessarily invalidate test scores. The following list of non-allowable accommodations was printed in all examiner manuals:

- Test administration in a language other than English, either orally or in writing
- Translation of assessment into any language other than English
- Translation of assessment into sign language
- Use of monolingual English dictionaries, bilingual dictionaries, or other similar comprehension aids
- Responses in native language

#### **4. IELA 2007 Test Security**

**4.1 Test Security Agreement.** All testing personnel as well as any individuals involved in transcriptions of student responses were required to sign a Test Security Agreement which included the following statements:

1. I will protect the contents of the test from any improper access.
2. I will handle test materials in accordance with security instructions. Copying or taking notes about any part of the test is not allowed.
3. I will carefully restrict access to the test materials to only persons authorized by the District Test Coordinator.
4. I will assure that students' responses are accurate reflections of their own work.
5. I will assure that students' answers to test items are their own and that no one offers any improper assistance to students.
6. I acknowledge that discussing with teachers or students or answering any test questions contained in the assessment before, during, or after the administration of the test is a violation of test security.

The District Test Coordinator was instructed (in the IELA 2008 Test Coordinator's Guide) to collect and file all signed copies of the Test Security Agreement.

**4.2 Bar-Coding and Return of Secure Materials.** All test booklets, prompt books, Listening test CDs, and examiner manuals were individually bar-coded. These secure test materials were scanned upon packing and distributing to districts and then scanned again upon return to Questar. Test Coordinators were instructed to return all test materials—used and unused—to Questar. Districts were informed of any materials missing from their return shipment.

**4.3 Storage and Shredding of Secure Materials.** After scoring, all used answer documents were transferred to secure storage facilities in Brewster, NY. Access to these facilities is limited to authorized personnel. Student answer documents will be stored for a minimum of 180 days, and then shredded.

Except for file copies, all unused and non-scannable secure test materials (examiner manuals, prompt books, and non-scannable test booklets) were shredded.

## 5. IELA 2007 Item Analyses

Item analyses included calculation of item difficulty and item discrimination. The typical classical item difficulty measure is the p-value statistic which is calculated as the proportion of students that responded correctly for each MC item or the average percent of the maximum score that students earned on each CR item. Thus p-value is inversely related to item difficulty (i.e., the lower the p-value, the more difficult the item). Table 2 shows the average p-value and range for IELA 2007 items by language modality and test form. Generally, it is desirable for tests to include items that span a range of difficulty. P-values are not sample independent measures of item difficulty, however. For example, the average p-value for items in each modality of Form B1 was lower than the average p-value in the corresponding modality on Form B2. This does not mean, however, that B1 included more difficult items than B2. This difference resulted from the differences in ability of the students who were administered these two forms. Generally, the students administered Form B1 were at lower levels of English proficiency than students who were administered Form B2. Item difficulty, as represented by p-values, can be used to determine the appropriateness of items for future versions of the IELA.

Point-biserial correlations are used to examine item discrimination. Point-biserial correlation coefficients range between -1.0 and +1.0 and indicate the extent to which an item differentiates between low-ability and high-ability students. High positive values indicate that a high-ability student is more likely to answer an item correctly and low negative values indicate that a low-ability student is more likely to answer an item correctly. Table 2 shows median point-biserial correlation coefficients and range by language modality and test form. Point-biserial correlations can also be used to determine the appropriateness of an item for future versions of the IELA. Analyses of test level data, including raw score descriptive statistics and test reliability measures, are reported in Table 6.

**Table 2. Summary of IELA 2007 Item Difficulty and Discrimination by Grade Cluster and Language Modality**

Grade Span	Form	Modality	Item p-value		Point Biserial	
			Average	Range	Median	Range
K	A	L	0.57	0.10 - 0.90	0.40	0.24 - 0.50
		S	0.68	0.32 - 0.88	0.45	0.29 - 0.57
		R	0.58	0.23 - 0.97	0.41	0.24 - 0.63
		W	0.59	0.26 - 0.97	0.46	0.19 - 0.56
1-2	B1	L	0.71	0.31 - 0.95	0.36	0.10 - 0.51
		S	0.66	0.46 - 0.88	0.65	0.51 - 0.72
		R	0.72	0.53 - 0.90	0.36	0.15 - 0.55
		W	0.62	0.27 - 0.92	0.55	0.28 - 0.68
	B2	L	0.83	0.44 - 0.99	0.29	0.19 - 0.39
		S	0.86	0.67 - 0.98	0.33	0.17 - 0.55
		R	0.79	0.52 - 0.96	0.34	0.13 - 0.42
		W	0.71	0.35 - 0.93	0.45	0.35 - 0.59
3-6	C1	L	0.59	0.35 - 0.84	0.40	0.15 - 0.61
		S	0.58	0.37 - 0.94	0.59	0.29 - 0.81
		R	0.56	0.31 - 0.85	0.36	0.12 - 0.54
		W	0.57	0.28 - 0.76	0.45	0.10 - 0.77
	C2	L	0.83	0.53 - 0.97	0.31	0.15 - 0.41
		S	0.92	0.80 - 1.00	0.24	0.0 - 0.51
		R	0.72	0.48 - 0.91	0.42	0.31 - 0.51
		W	0.75	0.38 - 0.94	0.37	0.29 - 0.58
7-8	D1	L	0.52	0.27 - 0.72	0.41	0.10 - 0.54
		S	0.49	0.28 - 0.85	0.64	0.43 - 0.76
		R	0.52	0.27 - 0.84	0.30	0.15 - 0.44
		W	0.62	0.33 - 0.83	0.47	0.16 - 0.67
	D2	L	0.85	0.52 - 0.96	0.33	0.14 - 0.43
		S	0.90	0.76 - 0.99	0.29	0.13 - 0.56
		R	0.68	0.32 - 0.89	0.38	0.19 - 0.54
		W	0.75	0.47 - 0.97	0.33	0.19 - 0.57

Grade Span	Form	Modality	Item p-value		Point Biserial	
9-12	E1	L	0.49	0.24 - 0.81	0.34	0.15 - 0.56
		S	0.42	0.18 - 0.72	0.46	0.32 - 0.72
		R	0.55	0.24 - 0.88	0.38	0.11 - 0.49
		W	0.40	0.20 - 0.73	0.38	0.08 - 0.69
	E2	L	0.81	0.63 - 0.97	0.37	0.17 - 0.52
		S	0.85	0.73 - 0.98	0.41	0.23 - 0.58
		R	0.74	0.30 - 0.96	0.39	0.16 - 0.56
		W	0.70	0.45 - 0.95	0.39	0.28 - 0.56

Analyses of test level data, including raw score descriptive statistics and test reliability measures, are reported in Table 5.

## 6. Scaling and Equating of the IELA

The IELA 2007 test forms were equated to IELA 2006 forms so that scores could be reported on the same score scale. Prior to equating 2007 to 2006 forms, however, Spring 2007 IELA items in each grade cluster test form were calibrated using the Rasch Partial Credit Model (PCM), as implemented in WINSTEPS, version 3.57.1. This model, which was used to calibrate IELA 2006 items, is appropriate for short-answer and constructed response items on the Speaking and Writing subtests as well as multiple-choice items administered across the language domains. As a first step, items on 2007 forms A, B2, C2, D2, and E2 were calibrated, with items on each grade-cluster form calibrated independently. Items on 2007 level 1 forms, B1, C1, D1, and E1, were then calibrated by fixing the item parameters for those items that are common between the two levels of each grade cluster (i.e., forms C1 and C2) to the same values as the level 2 calibration for those items. This calibration procedure equated Forms B1, C1, D1, and E1 to Forms B2, C2, D2, and E2, respectively, ensuring that, within each grade cluster, scores on the level 1 and level 2 forms are reported on the same scale.

Following the item calibration, IELA 2007 test forms were equated to the 2006 forms using a common item or anchor test design. Anchor items, those items that appeared in identical format in both the Spring 2006 form and in the Spring 2007 form, were embedded in Forms A, B2, C2, D2, and E2. Within each grade cluster, at least 30% of the items were in common between the 2006 and 2007 forms. Prior to equating 2007 to 2006 forms, each anchor item was evaluated for stability. As part of that evaluation, the calibrated difficulty (step value) of each anchor item in the current year (2007) was plotted against the calibrated difficulty of that item in the prior year (2006). Ideally, these plots should fall on a 45-degree line, indicating that calibrated values are stable from year to year. Those points that fall quite far from the line are referred to as outliers. For the anchor items in each of the five forms, the 2007 step values were plotted against the 2006 step values and these plots are shown in Figures 1 – 5. The numbers of plotted points for Forms A, B2, C2, D2, and E2 are 51, 35, 27, 31, and 27, respectively. The plots show that the step values fall along this 45 degree line as the model requires. Of course, not all points are on or right next to the line due to error that is inherent in all measurement, and occasionally, a point is quite far from the line. Across the five forms, there were only a few outliers and these outliers were removed from the equating. Once the items are initially equated, a difference is calculated between the two step values (2006 step value – 2007 step value). Outliers were defined as items with an absolute difference of 0.60 logits or greater. These items are represented by the points that are furthest from the 45-degree line in the plots. The items that were not used as part of the equating were still scored and used as operational items on their respective forms, but those items were not used in the calculations to determine final equating constants. (Note that when a constructed response item with multiple score points had at least one outlier point, the entire item was removed

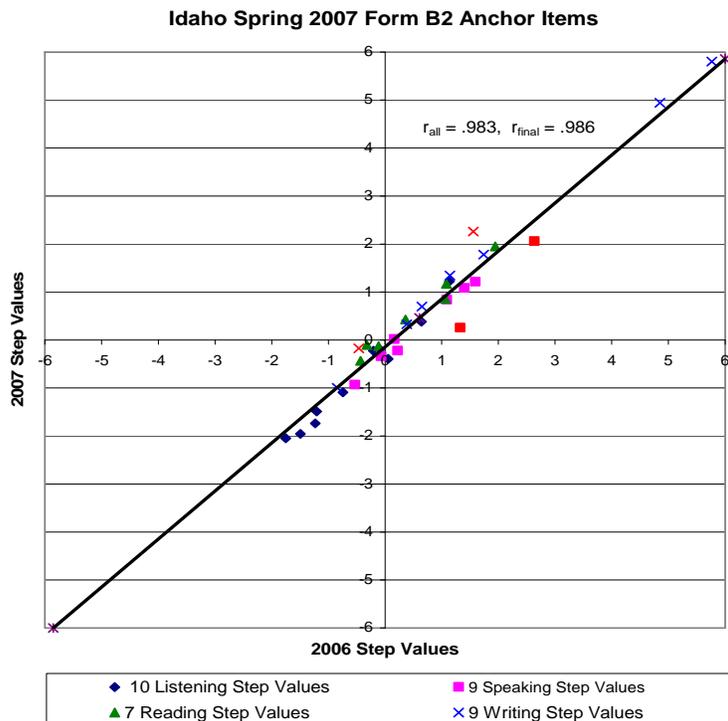
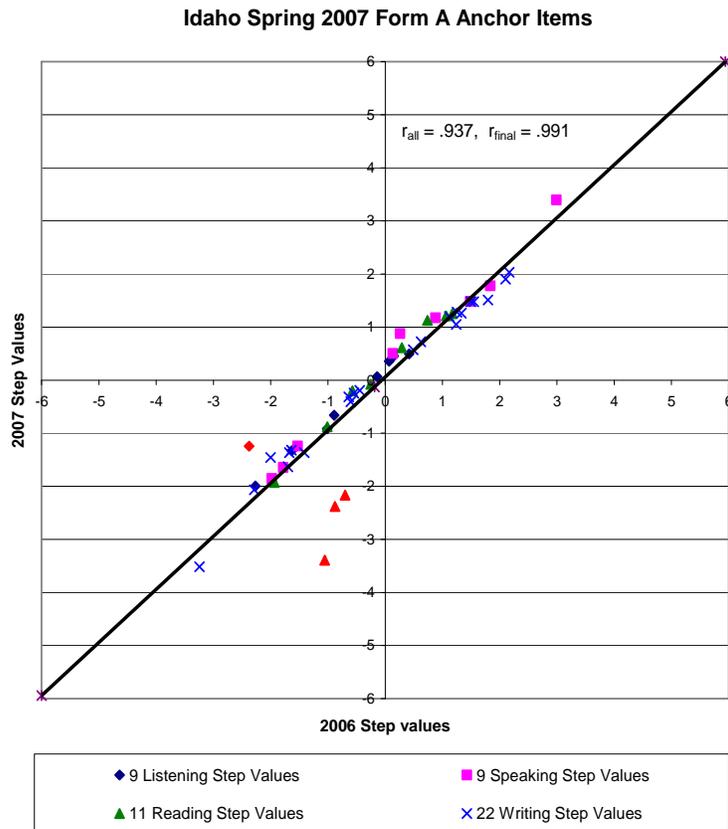
from the equating.) After deleting items with outlier values, the number of step values for the forms as listed previously is 47, 27, 26, 26, and 25. Table 3 shows the number of points represented by anchor items by form and modality both before (B) and after (A) outliers were removed.

**Table 3. Anchor Item Points (Step values) by Form and Modality**

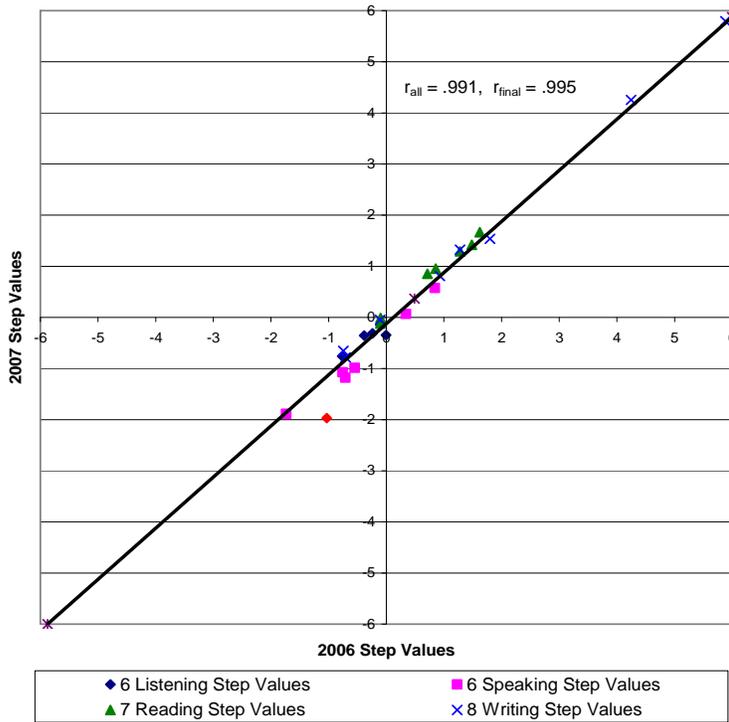
Form	Listening		Speaking		Reading		Writing		Total	
	B	A	B	A	B	A	B	A	B	A
<b>A</b>	9	8	9	9	11	8	22	22	51	47
<b>B2</b>	10	10	9	5	7	7	9	5	35	27
<b>C2</b>	6	5	6	6	7	7	8	8	27	26
<b>D2</b>	6	6	9	4	8	8	8	8	31	26
<b>E2</b>	7	7	5	3	8	8	7	7	27	25

In Figures 1 through 5, two correlation coefficients ( $r$ ) are given in the upper right-hand corner of each plot: one for all anchor items and the other for the final anchor items with outliers removed.

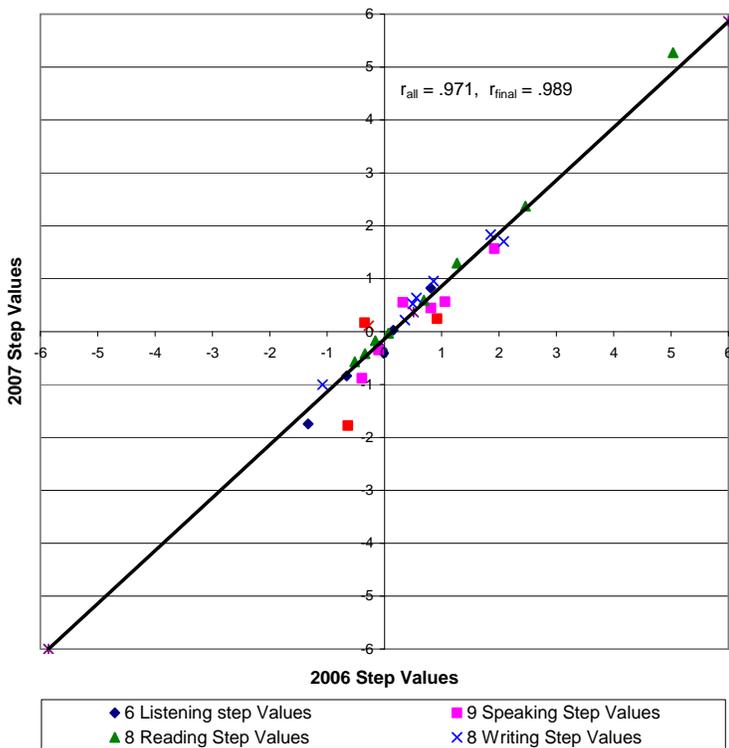
Figures 1-5. Step values of Anchor Items for 2006 and 2007 IELA forms.

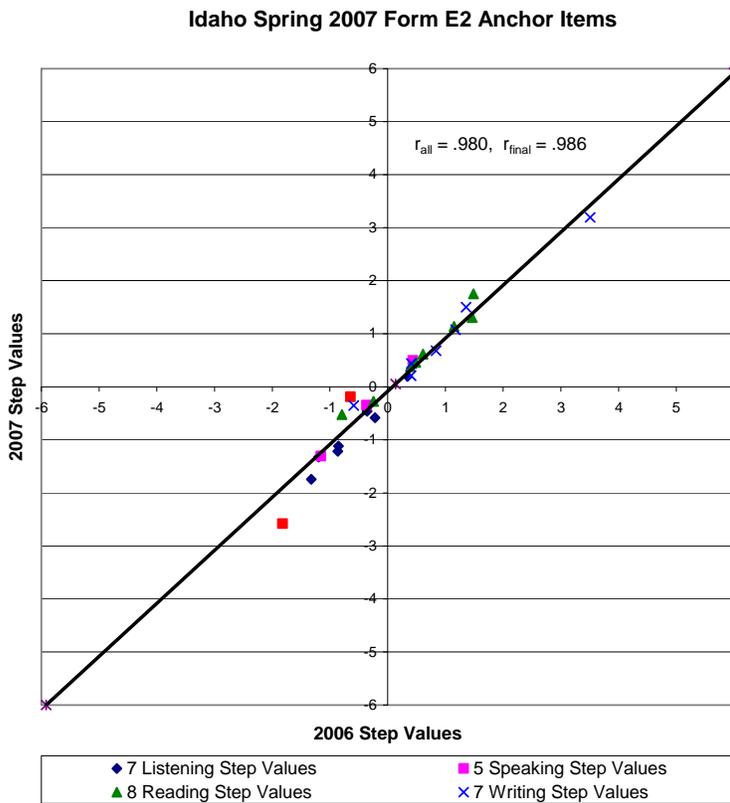


Idaho Spring 2007 Form C2 Anchor Items



Idaho Spring 2007 Form D2 Anchor Items





With the outliers removed, the final anchor items were used to develop a linking constant for each form that places the item step values from the 2007 form on the same Rasch logit scale as the 2006 form. The linking constant is computed as the difference between the average step value from the 2006 form’s Winsteps calibration, minus the average step value from the 2007 form’s Winsteps calibration. Adding this linking constant to the step values for each of the items in the 2007 form places all of the 2007 form’s step values (and log ability estimates) on the same Rasch logit scale as the 2006 form. A separate linking constant was calculated for each grade cluster and applied. This constant was applied to items on all forms including both the level 1 and level 2 forms.

Once all items from the 2006 and 2007 forms were placed on the same logit difficulty scale established in 2006, scale scores were computed for the 2007 forms. For the Total, scale scores were developed in 2006 for each grade cluster form by setting the Early Fluent and Fluent proficiency level cut scores to pre-specified values. For each subtest (L, S, R, W, C), scale scores were developed by setting the Advanced Beginning and Early Fluent proficiency level cuts to pre-specified values. The same linear transformation that was developed in the first year for each IELA 2006 grade cluster form and test was then applied to the equated Rasch log ability scale for the 2007 grade cluster form to yield equated scale scores.

Table 4 shows the number of items and number of step values that were deleted to yield the final anchor item equating. Equating constants were calculated both with the outliers included and with them deleted. The table also shows the effect on the equating of deleting the outliers by comparing the two sets of calculations. The effect is given in three different metrics, in terms of the change in scale scores, raw scores, and conditional standard error of measurement (SEM). The change in conditional SEM is done at the Early Fluent cut score which is a scale score with the smallest conditional SEM, thus the change in scale score as a percentage of SEM would be highest at this point. The change in raw score represents, in raw score units, the change in scale score over the range of scores from one SEM above to one SEM below the Early Fluent cut score. This is the point in the conversion tables where differences between scale scores for adjacent raw scores are the smallest. Across all five forms, the effect of deleting outliers on equating is small, if not trivial. The largest effect is for Form A with a 2.5 scale score change. However, the 2.5 scale score change represents only 1.5 raw scores on a 94 point assessment and is only one-third of the conditional SEM. In fact, these are the largest differences for this set of forms. For Form B2, there is no change by deleting the outliers. For the other three forms, the change ranges from .1 to .7 of a raw score and from 2% to 20% of the conditional SEM. Since the SEM represents variability in scores that could be attributed to error, the effect of removing the items from the equating was quite small. If the effect on raw scores or conditional SEM was evaluated at scale scores further from the Early Fluent cut score, then the change in raw score and conditional SEM would be even smaller. Overall, equating with the Rasch model via the anchor test design worked extremely well. Across the five forms, there were only a few discrepant points, and all the remaining points in each of the five plots were on or right next to the 45 degree line yielding correlations of .99. Even deleting the few outliers had, at most, only a small effect on the equating results.

**Table 4. Effect on Equating by Deleting Outlier Anchor Items**

<b>Spring 2007 Idaho English Language Assessment</b>					
			<b>Change in Scale Score</b>	<b>Change at the Early Fluent Cut Score in</b>	
<b>Form</b>	<b>Deleted</b>			<b>Raw Score</b>	<b>% Standard Error</b>
	<b># Items</b>	<b># Steps</b>			
<b>A</b>	4	4	-2.5	1.5	36
<b>B2</b>	2	8	0.0	0.0	0
<b>C2</b>	1	1	-0.7	0.5	11
<b>D2</b>	2	5	-1.0	0.7	20
<b>E2</b>	2	2	-0.1	0.1	2

## 7. Reliability of the IELA 2007

Data bearing on the reliability of IELA 2007 Test Forms are shown in the panels of Table 5. The tables show for each form and each language domain (and comprehension and the total test) the number of students (N) who were administered the form, coefficient Alpha, a measure of internal-consistency reliability, the maximum raw score attainable, and the mean, standard deviation, and standard error of measurement (SEM) in both raw score and scale score units. This table includes scores for students identified as LEP (limited English proficient) and LEP1<sup>1</sup> but not those identified as LEPX<sup>2</sup>. Number of students represents the number for whom there was a valid test score and may vary across language domains in a grade to the extent that there were students who did not attempt one or more of the language domain tests. There is a total score for each student regardless of whether or not all language domain tests were attempted.

---

<sup>1</sup> New to U.S. school within the last 12 months.

<sup>2</sup> Exited out of an LEP program within the last 2 years and not required to test on the IELA.

**Table 5. Reliability, Raw Score and Scale Score Descriptive Statistics for IELA 2007 Test Forms by Grade**

Grade K				Raw Scores				Scale Scores		
Form	Language Domain	N	Alpha	Max	Mean	Std. Dev.	SEM	Mean	Std. Dev.	SEM
<b>A</b>	Listening	2,397	0.86	22	11.4	4.5	1.71	101.6	22.7	8.65
	Speaking	2,384	0.84	22	12.9	5.2	2.07	101.3	24.1	9.63
	Reading	2,393	0.92	36	17.5	7.1	1.99	101.6	25.6	7.19
	Writing	2,331	0.94	22	13.0	6.3	1.49	104.2	34.4	8.13
	Comprehen	2,402	0.87	29	12.3	5.4	1.96	100.8	21.7	7.83
	Total	2,407	0.95	102	54.1	18.4	4.02	400.6	35.8	7.84

Grade 1				Raw Scores				Scale Scores		
Form	Language Domain	N	Alpha	Max	Mean	Std. Dev.	SEM	Mean	Std. Dev.	SEM
<b>B1</b>	Listening	216	0.78	22	15.3	3.7	1.72	100.3	15.7	7.29
	Speaking	211	0.90	22	13.5	6.2	2.01	97.4	28.6	9.27
	Reading	217	0.73	15	10.7	2.9	1.52	97.5	17.2	8.99
	Writing	217	0.86	15	8.9	3.9	1.47	97.0	23.5	8.94
	Comprehen	217	0.80	31	21.5	4.9	2.18	98.1	14.0	6.21
	Total	218	0.93	74	47.8	14.5	3.78	391.8	45.6	11.89

<b>B2</b>	Listening	1,758	0.71	22	17.2	2.9	1.58	109.3	15.9	8.55
	Speaking	1,754	0.79	22	16.6	4.2	1.95	111.2	19.4	8.99
	Reading	1,758	0.73	20	14.3	3.3	1.70	107.8	15.7	8.21
	Writing	1,758	0.81	20	10.3	3.6	1.59	106.2	18.5	8.13
	Comprehen	1,759	0.81	39	29.0	5.2	2.27	107.7	13.7	6.01
	Total	1,765	0.90	84	58.1	11.6	3.68	419.0	37.3	11.80

Grade 2				Raw Scores				Scale Scores		
Form	Language Domain	N	Alpha	Max	Mean	Std. Dev.	SEM	Mean	Std. Dev.	SEM
<b>B1</b>	Listening	107	0.83	22	15.6	4.1	1.68	103.0	20.7	8.53
	Speaking	107	0.91	22	12.2	7.1	2.11	91.4	33.9	10.05
	Reading	106	0.84	15	10.8	3.5	1.37	100.3	21.8	8.59
	Writing	107	0.91	15	9.1	4.7	1.43	99.9	29.6	9.07
	Comprehen	107	0.88	31	22.1	5.8	2.03	101.8	19.2	6.68
	Total	108	0.96	74	47.1	18.0	3.65	394.8	59.5	12.04

<b>B2</b>	Listening	1,560	0.72	22	19.3	2.2	1.18	122.0	16.8	8.96
	Speaking	1,558	0.77	22	18.8	3.3	1.62	123.3	19.8	9.60
	Reading	1,563	0.71	20	17.1	2.5	1.35	123.7	17.0	9.16
	Writing	1,562	0.73	20	13.6	2.7	1.42	124.3	16.8	8.75
	Comprehen	1,563	0.81	39	33.7	4.1	1.77	122.3	15.6	6.79
	Total	1,567	0.89	84	68.5	9.2	3.11	457.3	38.8	13.09

Grade 3				Raw Scores				Scale Scores		
Form	Language Domain	N	Alpha	Max	Mean	Std. Dev.	SEM	Mean	Std. Dev.	SEM
C1	Listening	84	0.87	22	12.2	5.2	1.89	91.6	15.3	5.51
	Speaking	82	0.91	22	11.0	6.8	2.04	83.1	25.2	7.59
	Reading	86	0.77	15	7.9	3.6	1.71	89.2	15.5	7.45
	Writing	85	0.82	15	6.7	3.7	1.57	85.4	18.2	7.82
	Comprehen	86	0.89	31	16.1	7.3	2.41	89.9	15.7	5.17
	Total	86	0.95	74	37.0	17.4	3.92	371.8	32.0	7.24
C2	Listening	1,399	0.78	22	16.7	3.7	1.76	105.2	13.1	6.19
	Speaking	1,398	0.74	22	18.4	3.3	1.71	107.3	15.0	7.71
	Reading	1,399	0.81	20	12.0	4.3	1.92	103.2	13.6	5.97
	Writing	1,397	0.76	19	10.7	3.3	1.65	103.8	13.3	6.55
	Comprehen	1,401	0.86	39	26.1	6.9	2.59	103.7	11.5	4.33
	Total	1,401	0.91	83	57.6	11.9	3.67	407.1	20.4	6.27

**Grade 4**

Form	Language Domain	N	Alpha	Max	Mean	Std. Dev.	SEM	Mean	Std. Dev.	SEM
C1	Listening	87	0.83	22	12.3	4.8	1.98	92.6	13.8	5.67
	Speaking	86	0.90	22	10.8	6.4	2.03	81.8	22.7	7.23
	Reading	88	0.73	15	8.1	3.3	1.69	89.6	14.2	7.41
	Writing	87	0.79	15	7.8	3.5	1.61	91.4	17.0	7.80
	Comprehen	88	0.87	30	16.2	6.8	2.45	90.7	14.5	5.23
	Total	88	0.94	74	38.5	16.4	3.87	375.2	29.2	6.90
C2	Listening	1,362	0.78	22	18.3	3.0	1.42	111.8	13.4	6.33
	Speaking	1,360	0.79	22	19.5	2.9	1.33	113.6	15.4	7.07
	Reading	1,370	0.81	20	14.5	4.1	1.76	111.9	15.1	6.52
	Writing	1,369	0.73	19	12.4	3.0	1.57	111.3	14.0	7.26
	Comprehen	1,370	0.86	39	30.1	6.2	2.34	111.2	12.9	4.87
	Total	1,371	0.91	83	64.5	11.1	3.35	420.8	23.0	6.95

**Grade 5**

Form	Language Domain	N	Alpha	Max	Mean	Std. Dev.	SEM	Mean	Std. Dev.	SEM
C1	Listening	87	0.86	22	14.0	5.1	1.88	97.4	15.3	5.67
	Speaking	88	0.89	22	12.0	6.8	2.22	85.4	24.4	8.02
	Reading	87	0.76	15	9.1	3.2	1.60	93.8	14.0	6.89
	Writing	87	0.80	15	8.1	3.6	1.61	91.6	17.4	7.75
	Comprehen	87	0.89	31	18.7	6.8	2.30	95.6	14.3	4.81
	Total	88	0.95	74	42.9	17.0	3.91	381.9	31.9	7.35
C2	Listening	1,210	0.76	22	19.1	2.6	1.30	115.5	13.5	6.65
	Speaking	1,208	0.78	22	19.9	2.6	1.21	116.1	14.9	6.96
	Reading	1,215	0.81	20	15.8	3.7	1.63	116.5	14.6	6.45
	Writing	1,215	0.74	19	13.3	2.9	1.45	115.8	14.4	7.33
	Comprehen	1,216	0.85	39	32.0	5.6	2.19	115.1	12.7	4.95
	Total	1,216	0.90	83	67.9	10.0	3.09	428.9	23.1	7.14

<b>Grade 6</b>				<b>Raw Scores</b>				<b>Scale Scores</b>		
<b>Form</b>	<b>Language Domain</b>	<b>N</b>	<b>Alpha</b>	<b>Max</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>SEM</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>SEM</b>
<b>D1</b>	Listening	71	0.82	22	11.9	4.7	2.00	87.3	10.8	4.62
	Speaking	69	0.91	22	9.9	7.0	2.12	79.9	20.1	6.14
	Reading	72	0.65	15	7.4	2.9	1.69	83.7	10.3	6.08
	Writing	72	0.85	15	7.8	3.9	1.51	84.3	15.9	6.26
	Comprehen	72	0.83	32	16.2	6.2	2.59	85.7	9.8	4.10
	Total	72	0.94	74	36.4	16.5	3.94	365.7	24.0	5.74
<b>D2</b>	Listening	1,120	0.76	22	18.0	3.1	1.53	102.3	10.1	4.91
	Speaking	1,120	0.77	22	18.5	3.6	1.72	103.8	12.2	5.82
	Reading	1,121	0.76	24	14.3	4.3	2.12	101.4	9.0	4.40
	Writing	1,119	0.75	20	12.6	3.2	1.61	101.8	10.3	5.16
	Comprehen	1,123	0.84	43	30.5	6.7	2.65	100.6	8.3	3.29
	Total	1,124	0.91	88	63.1	12.1	3.72	402.3	17.2	5.30

**Grade 7**

<b>Form</b>	<b>Language Domain</b>	<b>N</b>	<b>Alpha</b>	<b>Max</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>SEM</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>SEM</b>
<b>D1</b>	Listening	82	0.85	22	11.4	5.1	2.01	86.7	11.9	4.67
	Speaking	77	0.92	22	10.0	7.1	1.97	80.7	21.1	5.81
	Reading	83	0.73	15	8.0	3.2	1.66	85.6	11.4	5.99
	Writing	83	0.79	15	8.0	3.7	1.68	85.7	14.5	6.62
	Comprehen	83	0.88	32	16.5	7.1	2.52	86.1	11.1	3.92
	Total	83	0.95	74	36.6	17.2	3.93	366.3	25.3	5.77
<b>D2</b>	Listening	1,092	0.75	22	18.6	3.0	1.49	104.6	10.3	5.19
	Speaking	1,090	0.76	22	19.1	3.3	1.62	106.4	12.3	6.02
	Reading	1,090	0.78	24	15.5	4.3	2.06	103.9	9.5	4.48
	Writing	1,091	0.72	20	13.2	3.1	1.62	104.0	10.2	5.41
	Comprehen	1,092	0.85	44	32.3	6.5	2.55	103.0	8.6	3.36
	Total	1,093	0.90	88	66.3	11.3	3.58	407.3	16.8	5.32

**Grade 8**

<b>Form</b>	<b>Language Domain</b>	<b>N</b>	<b>Alpha</b>	<b>Max</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>SEM</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>SEM</b>
<b>D1</b>	Listening	78	0.79	22	10.1	4.6	2.11	84.2	10.9	5.00
	Speaking	71	0.91	22	7.8	6.1	1.88	73.8	19.1	5.86
	Reading	78	0.73	15	7.1	3.2	1.65	82.4	12.1	6.29
	Writing	78	0.82	15	7.8	3.6	1.52	84.3	14.2	5.94
	Comprehen	78	0.84	32	14.5	6.4	2.56	83.4	9.7	3.87
	Total	78	0.94	74	32.1	15.7	3.79	359.9	22.8	5.49
<b>D2</b>	Listening	960	0.80	22	18.8	3.1	1.38	105.5	10.8	4.88
	Speaking	959	0.81	22	19.1	3.5	1.54	106.7	12.9	5.70
	Reading	961	0.76	24	16.0	4.2	2.05	105.0	9.3	4.50
	Writing	962	0.72	20	13.5	3.0	1.62	104.7	10.4	5.52
	Comprehen	963	0.84	44	32.9	6.4	2.54	103.9	8.6	3.41
	Total	963	0.90	88	67.2	11.3	3.53	408.9	17.2	5.35

Grade 9				Raw Scores				Scale Scores		
Form	Language Domain	N	Alpha	Max	Mean	Std. Dev.	SEM	Mean	Std. Dev.	SEM
E1	Listening	83	0.78	22	9.9	4.4	2.05	84.7	9.9	4.66
	Speaking	80	0.89	22	7.2	5.8	1.91	74.7	17.4	5.75
	Reading	84	0.75	15	7.4	3.4	1.69	81.4	11.8	5.91
	Writing	84	0.75	15	4.8	3.1	1.55	80.8	12.6	6.24
	Comprehen	84	0.84	32	14.1	6.3	2.54	82.8	9.6	3.88
	Total	84	0.93	74	28.8	14.3	3.82	366.0	16.4	4.38
E2	Listening	903	0.81	22	17.2	3.7	1.61	101.6	10.9	4.72
	Speaking	904	0.82	22	17.3	4.3	1.80	101.6	12.8	5.41
	Reading	903	0.81	25	15.8	4.5	1.94	99.7	10.0	4.33
	Writing	900	0.77	20	11.7	3.4	1.66	100.2	10.1	4.90
	Comprehen	908	0.88	44	31.3	7.6	2.60	99.8	9.5	3.26
	Total	910	0.93	89	61.5	14.0	3.82	399.2	15.2	4.17

**Grade 10**

Form	Language Domain	N	Alpha	Max	Mean	Std. Dev.	SEM	Mean	Std. Dev.	SEM
E1	Listening	70	0.79	22	11.2	4.5	2.09	87.2	10.1	4.68
	Speaking	70	0.87	22	9.3	6.2	2.19	81.5	16.4	5.83
	Reading	70	0.69	15	8.6	3.0	1.67	85.4	10.0	5.54
	Writing	70	0.74	15	5.9	3.0	1.54	85.1	11.3	5.75
	Comprehen	70	0.85	32	16.5	6.4	2.51	85.9	9.2	3.61
	Total	70	0.92	74	35.0	14.2	3.93	373.0	15.7	4.34
E2	Listening	782	0.85	22	17.7	3.8	1.48	103.5	12.2	4.68
	Speaking	783	0.85	22	17.9	4.3	1.67	104.5	13.9	5.35
	Reading	788	0.81	25	16.8	4.5	1.94	102.0	10.3	4.47
	Writing	785	0.78	20	12.1	3.7	1.71	101.8	11.3	5.28
	Comprehen	790	0.90	45	32.6	7.8	2.52	101.9	10.3	3.33
	Total	792	0.94	89	63.9	14.7	3.75	402.7	17.1	4.36

**Grade 11**

Form	Language Domain	N	Alpha	Max	Mean	Std. Dev.	SEM	Mean	Std. Dev.	SEM
E1	Listening	36	0.80	22	11.7	4.6	2.09	88.0	9.2	4.18
	Speaking	36	0.84	22	9.4	5.5	2.18	80.8	15.1	5.99
	Reading	36	0.77	15	9.1	3.4	1.61	87.3	12.7	6.01
	Writing	36	0.78	15	6.1	3.3	1.53	85.8	11.6	5.45
	Comprehen	36	0.84	32	17.5	6.3	2.50	87.4	8.9	3.51
	Total	36	0.92	74	36.2	14.0	3.89	374.3	15.0	4.17
E2	Listening	644	0.81	22	18.1	3.5	1.52	104.2	10.9	4.75
	Speaking	640	0.85	22	18.2	4.0	1.59	105.3	13.3	5.23
	Reading	642	0.81	25	17.4	4.2	1.86	103.4	10.2	4.49
	Writing	640	0.78	20	12.6	3.4	1.63	103.0	10.9	5.18
	Comprehen	647	0.87	45	33.5	7.1	2.54	103.0	9.6	3.43
	Total	647	0.92	89	65.6	13.3	3.67	404.5	15.9	4.38

Grade 12				Raw Scores				Scale Scores		
Form	Language Domain	N	Alpha	Max	Mean	Std. Dev.	SEM	Mean	Std. Dev.	SEM
E1	Listening	20	0.88	22	11.2	5.8	1.96	87.3	13.6	4.65
	Speaking	20	0.77	22	10.6	4.9	2.33	85.2	10.7	5.11
	Reading	20	0.75	15	9.1	3.4	1.67	87.7	13.0	6.47
	Writing	20	0.59	15	6.1	2.5	1.61	85.8	8.3	5.34
	Comprehen	20	0.88	32	16.8	7.3	2.47	86.5	10.7	3.64
	Total	20	0.93	74	37.0	14.3	3.89	375.2	15.5	4.23
E2	Listening	399	0.87	22	17.7	3.8	1.36	103.1	11.5	4.14
	Speaking	401	0.86	22	18.4	3.9	1.44	106.2	13.5	4.98
	Reading	401	0.84	25	17.4	4.4	1.74	103.5	10.3	4.09
	Writing	399	0.79	20	12.6	3.4	1.59	103.3	11.0	5.06
	Comprehen	402	0.91	45	33.4	7.6	2.23	102.8	10.0	2.95
	Total	408	0.94	89	64.9	14.6	3.70	403.6	17.5	4.43

## 8. Validity of the IELA

**8.1 Content and Construct Validity.** Validity of the IELA begins with test content. The Introduction to the Mountain West Assessment Consortium Foundation Document, included as an Appendix, provides background information on the design of the assessment.

In addition to test design considerations, test results also bear on the content validity of the assessment. In very general terms, the distribution and range of scores within each grade cluster and grade level (Table 5) provide evidence that the IELA can capture a range of abilities. And, Table 6 provides information on the validity of the assessment showing intercorrelations among components of the test. This table shows, by grade cluster and by test form, Pearson product moment correlations among scale scores on each subtest (Listening, Speaking, Reading, Writing, and Comprehension). Correlations are not reported for subtests that share common items (e.g., Reading and Comprehension) nor are they reported for subtests and Total IELA. The number below the correlation coefficient in each cell represents the number of students on which the correlation is based.

**Table 6. Correlations Among Scale Scores on Individual Language Domain Tests**

Grade	K	1-2		3-5		6-8		9-12		
r	A	B1	B2	C1	C2	D1	D2	E1	E2	Avg.
<b>L x S</b>	0.71 2,383	0.64 318	0.47 3,555	0.66 255	0.42 4,486	0.70 218	0.43 3,654	0.52 205	0.48 3,032	0.56
<b>L x R</b>	0.55 2,391	0.57 324	0.59 3,564	0.73 258	0.58 4,492	0.67 232	0.58 3,656	0.61 209	0.62 3,038	0.61
<b>L x W</b>	0.33 2,325	0.6 324	0.55 3,563	0.66 257	0.52 4,490	0.72 232	0.47 3,655	0.56 209	0.57 3,031	0.55
<b>S x R</b>	0.54 2,386	0.60 318	0.42 3,556	0.57 255	0.41 4,486	0.57 218	0.43 3,654	0.60 206	0.47 3,033	0.51
<b>S x W</b>	0.35 2,312	0.67 319	0.46 3,554	0.71 255	0.41 4,484	0.66 218	0.37 3,653	0.53 206	0.4 3,026	0.51
<b>S x C</b>	0.70 2,387	0.64 319	0.49 3,556	0.66 255	0.45 4,489	0.69 218	0.48 3,659	0.63 206	0.52 3,043	0.58
<b>R x W</b>	0.46 2,320	0.67 324	0.66 3,567	0.68 259	0.66 4,505	0.62 234	0.65 3,659	0.67 210	0.70 3,048	0.64
<b>W x C</b>	0.37 2,329	0.68 325	0.66 3,568	0.69 259	0.65 4,506	0.72 234	0.65 3,662	0.68 210	0.71 3,050	0.65
<b>Avg.</b>	0.50	0.63	0.54	0.67	0.51	0.67	0.51	0.60	0.56	0.58

All of the correlation coefficients in Table 6 are significantly different from zero, indicating that the different subtests are measuring related abilities. Insofar as the language domain tests are measuring aspects of the same construct, English proficiency, performance in the different domains should be related. In addition, however, the coefficients are not high enough to suggest that the abilities measured by the individual domain tests are identical, reinforcing the conclusion that language domain abilities are different aspects of overall English proficiency.

**8.2 Criterion-related Validity.** The performance of different subpopulations of LEP students also bears on the validity of the assessment. Table 7 shows, for each grade cluster and LEP group, the number of students to whom the test was administered (N) and mean and standard deviation of the scale scores for each language domain plus comprehension and the total test. These data are collapsed over grades and test forms within a grade cluster. Several points can be made from reviewing this table. First, for each grade cluster, a large majority of students who were administered the IELA were in the LEP rather than LEP1 or LEPX group. The proportion of LEP1 students was higher in Kindergarten than in other grade clusters. Second, in each grade cluster and for each language domain test and the total test, scores for LEPX students were higher on average than either LEP or LEP1. This difference was smaller in the higher grades, i.e., middle and high school, than in the lower grades. Third, for all grade clusters except K, scores for LEP1 students were lower on average than those of LEP

students. Because LEP status (LEP1, LEP, LEPX) was determined independently of scores on this test and is based on criteria related to English proficiency, the differences in scores by LEP condition can be used as a source of criterion-related validity. All of these findings are consistent with results on the 2006 IELA.

**Table 7. LEP Groups Scale Scores by Grade Cluster**

<b>LEP1</b>				<b>LEP</b>			<b>LEPX</b>		
<b>IELA-A</b>	<b>N</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>N</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>N</b>	<b>Mean</b>	<b>Std. Dev.</b>
Listening	687	99.1	24.1	1,711	102.7	22.1	3	138.0	28.4
Speaking	681	98.7	25.5	1,704	102.3	23.4	3	136.7	27.7
Reading	685	100.0	26.5	1,709	102.3	25.2	3	132.0	15.6
Writing	684	102.0	35.7	1,648	105.1	33.8	3	129.0	27.7
Comprehen	689	98.6	22.5	1,714	101.7	21.4	3	131.3	5.8
<b>Total</b>	<b>690</b>	<b>397.2</b>	<b>38.1</b>	<b>1,718</b>	<b>402.0</b>	<b>34.8</b>	<b>3</b>	<b>454.7</b>	<b>22.4</b>

<b>IELA-B</b>	<b>N</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>N</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>N</b>	<b>Mean</b>	<b>Std. Dev.</b>
Listening	215	104.3	19.1	3,430	114.6	17.7	248	123.9	17.2
Speaking	214	97.4	30.1	3,420	116.1	21.4	246	123.2	19.7
Reading	215	104.0	21.1	3,433	114.4	18.5	248	123.9	18.7
Writing	216	103.6	27.4	3,432	113.8	20.4	248	125.1	20.1
Comprehen	215	104.0	18.0	3,435	113.8	16.7	249	122.6	16.9
<b>Total</b>	<b>217</b>	<b>402.8</b>	<b>56.9</b>	<b>3,445</b>	<b>435.0</b>	<b>43.5</b>	<b>251</b>	<b>457.1</b>	<b>52.6</b>

<b>IELA-C</b>	<b>N</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>N</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>N</b>	<b>Mean</b>	<b>Std. Dev.</b>
Listening	232	96.8	16.9	3,999	110.4	14.1	524	118.3	13.6
Speaking	234	87.9	26.9	3,990	111.7	15.9	524	117.4	14.7
Reading	232	95.5	17.0	4,015	109.9	15.7	525	119.9	14.5
Writing	232	93.5	18.7	4,010	109.7	15.1	525	119.3	13.9
Comprehen	233	95.7	16.3	4,017	109.4	13.5	525	118.6	12.9
<b>Total</b>	<b>235</b>	<b>383.7</b>	<b>35.8</b>	<b>4,017</b>	<b>417.7</b>	<b>24.5</b>	<b>525</b>	<b>434.8</b>	<b>22.1</b>

<b>IELA-D</b>	<b>N</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>N</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>N</b>	<b>Mean</b>	<b>Std. Dev.</b>
Listening	204	87.8	13.0	3,200	103.8	10.7	491	109.4	9.2
Speaking	191	79.5	21.5	3,196	105.3	12.9	491	110.8	11.3
Reading	204	86.0	12.4	3,202	103.0	9.7	492	109.3	8.8
Writing	204	86.3	15.8	3,202	103.2	10.6	492	110.1	10.0
Comprehen	205	86.7	11.4	3,207	102.2	8.9	492	108.2	8.0
<b>Total</b>	<b>206</b>	<b>366.5</b>	<b>28.1</b>	<b>3,208</b>	<b>405.5</b>	<b>17.8</b>	<b>492</b>	<b>418.5</b>	<b>15.4</b>

<b>IELA-E</b>	<b>N</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>N</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>N</b>	<b>Mean</b>	<b>Std. Dev.</b>
Listening	190	88.6	11.6	2,747	102.7	11.6	324	109.2	11.2
Speaking	187	81.5	18.1	2,747	103.6	13.8	324	110.6	11.1
Reading	191	86.8	12.9	2,753	101.5	10.6	326	108.3	9.6
Writing	191	85.8	13.3	2,743	101.5	11.1	326	108.8	9.3
Comprehen	191	87.1	11.2	2,766	101.3	10.1	327	108.1	9.4
<b>Total</b>	<b>191</b>	<b>374.3</b>	<b>19.2</b>	<b>2,776</b>	<b>401.6</b>	<b>16.9</b>	<b>327</b>	<b>413.5</b>	<b>13.5</b>

## 9. IELA Performance by Year

Table 8 shows results for both 2006 and 2007 by form and grade, thus allowing a comparison of performance in those two years. This table shows, for each language domain, comprehension, and total IELA, the maximum raw score ( $RS_{Max}$ ), number of students ( $N$ ) administered the assessment, the average raw score ( $RS_{Mean}$ ) and average scale score ( $SS_{Mean}$ ). The table includes data for students classified as LEP and LEP1 but not LEPX. There are several generalizations that can be made. First, although the numbers of students tested per grade was relatively stable from 2006 to 2007, the numbers and percent of students who were administered level 1 forms (i.e., B1, C1, D1, E1) was lower in 2007 than in 2006 and there was a concomitant increase in the numbers and percent administered level 2 forms. This shift was a likely result of a change in policy to restrict the use of level 1 forms to those students who are considered LEP1. There was a second notable change from 2006 to 2007 that resulted from the change in distribution of level 1 and level 2 form administrations. In most grades, there was a decline in average performance on level 1 forms, but a smaller amount of change on level 2 forms. Because administration of level 1 forms in 2006 was less restricted, those forms were not ability appropriate for many of the students to whom they were administered, thus inflating 2006 scores. The change in policy thus provided for more appropriate use of level 1 forms. Concerning those grades in which there was a change in performance on level 2 forms, because this is not a matched sample, it is not possible to infer from these data that the level of English proficiency for individual students has increased or decreased. Student migration in and out of the pool can account in part for the results. Those students who are new to the program would generally be operating at lower levels of English proficiency; whereas those students who are being reclassified as LEPX would be those at the higher levels of proficiency. This inward and outward migration could make it appear as though there was an overall decline in ability. Third, in most grades on the level 2 forms, average raw scores on both listening and speaking tests were very high, relative to the maximum raw score, in both 2006 and 2007. This finding will be addressed when new forms are configured for the 2008 IELA administration.

**Table 8. Performance on 2006 and 2007 IELA Test Forms by Grade**

<b>Kindergarten</b>			<b>2006</b>			<b>2007</b>		
<b>Form</b>	<b>Language Domain</b>	<b>RS<sub>Max</sub></b>	<b>N</b>	<b>RS<sub>Mean</sub></b>	<b>SS<sub>Mean</sub></b>	<b>N</b>	<b>RS<sub>Mean</sub></b>	<b>SS<sub>Mean</sub></b>
<b>A</b>	Listening	22	2,054	12.5	101.1	2,397	11.4	101.6
	Speaking	22	2,046	12.7	100.9	2,384	12.9	101.3
	Reading	36	2,031	19.8	100.6	2,393	17.5	101.6
	Writing	22	1,983	12.8	104.1	2,331	13.0	104.2
	Comprehen	29	2,059	14.3	100.5	2,402	12.3	100.8
	<b>Total</b>	<b>102</b>	<b>2,071</b>	<b>56.7</b>	<b>399.7</b>	<b>2,407</b>	<b>54.1</b>	<b>400.6</b>
<b>Grade 1</b>			<b>2006</b>			<b>2007</b>		
<b>B1</b>	Listening	22	658	16.0	103.0	216	15.3	100.3
	Speaking	22	657	13.4	101.3	211	13.5	97.4
	Reading	15	651	11.2	100.2	217	10.7	97.5
	Writing	15	650	9.8	101.7	217	8.9	97.0
	Comprehen	31	666	22.2	100.3	217	21.5	98.1
	<b>Total</b>	<b>74</b>	<b>667</b>	<b>49.5</b>	<b>399.8</b>	<b>218</b>	<b>47.8</b>	<b>391.8</b>
<b>B2</b>	Listening	22	1,188	17.4	109.2	1,758	17.2	109.3
	Speaking	22	1,183	16.2	112.9	1,754	16.6	111.2
	Reading	20	1,225	15.0	107.8	1,758	14.3	107.8
	Writing	20	1,223	11.0	105.4	1,758	10.3	106.2
	Comprehen	39	1,227	29.3	106.7	1,759	29.0	107.7
	<b>Total</b>	<b>84</b>	<b>1,227</b>	<b>58.4</b>	<b>418.1</b>	<b>1,765</b>	<b>58.1</b>	<b>419.0</b>
<b>Grade 2</b>			<b>2006</b>			<b>2007</b>		
<b>B1</b>	Listening	22	406	17.6	111.8	107	15.6	103.0
	Speaking	22	402	15.4	110.4	107	12.2	91.4
	Reading	15	403	12.5	111.4	106	10.8	100.3
	Writing	15	396	11.4	112.5	107	9.1	99.9
	Comprehen	31	411	24.8	110.3	107	22.1	101.8
	<b>Total</b>	<b>74</b>	<b>411</b>	<b>55.7</b>	<b>423.5</b>	<b>108</b>	<b>47.1</b>	<b>394.8</b>
<b>B2</b>	Listening	22	1,283	19.2	120.0	1,560	19.3	122.0
	Speaking	22	1,281	18.1	121.8	1,558	18.8	123.3
	Reading	20	1,300	17.2	121.3	1,563	17.1	123.7
	Writing	20	1,298	14.3	123.6	1,562	13.6	124.3
	Comprehen	39	1,300	33.4	119.2	1,563	33.7	122.3
	<b>Total</b>	<b>84</b>	<b>1,300</b>	<b>68.3</b>	<b>452.2</b>	<b>1,567</b>	<b>68.5</b>	<b>457.3</b>

<b>Grade 3</b>			<b>2006</b>			<b>2007</b>		
<b>Form</b>	<b>Language Domain</b>	<b>RS<sub>Max</sub></b>	<b>N</b>	<b>RS<sub>Mean</sub></b>	<b>SS<sub>Mean</sub></b>	<b>N</b>	<b>RS<sub>Mean</sub></b>	<b>SS<sub>Mean</sub></b>
<b>C1</b>	Listening	22	398	14.6	99.2	84	12.2	91.6
	Speaking	22	398	15.4	98.5	82	11.0	83.1
	Reading	15	396	9.9	99.8	86	7.9	89.2
	Writing	15	397	10.0	99.5	85	6.7	85.4
	Comprehen	31	399	19.7	98.7	86	16.1	89.9
	Total	74	399	49.8	395.9	86	37.0	371.8
<b>C2</b>	Listening	22	1,230	17.0	106.6	1,399	16.7	105.2
	Speaking	22	1,228	18.4	109.6	1,398	18.4	107.3
	Reading	20	1,234	13.3	105.5	1,399	12.0	103.2
	Writing	19	1,231	12.0	105.5	1,397	10.7	103.8
	Comprehen	39	1,234	27.5	105.4	1,401	26.1	103.7
	Total	83	1,234	60.5	411.0	1,401	57.6	407.1
<b>Grade 4</b>			<b>2006</b>			<b>2007</b>		
<b>C1</b>	Listening	22	350	15.5	102.0	87	12.3	92.6
	Speaking	22	347	16.4	103.0	86	10.8	81.8
	Reading	15	350	10.6	102.4	88	8.1	89.6
	Writing	15	348	10.5	102.7	87	7.8	91.4
	Comprehen	30	351	21.1	101.6	88	16.2	90.7
	Total	74	351	52.6	402.0	88	38.5	375.2
<b>C2</b>	Listening	22	1,169	18.2	111.6	1,362	18.3	111.8
	Speaking	22	1,171	19.4	114.4	1,360	19.5	113.6
	Reading	20	1,168	14.9	111.2	1,370	14.5	111.9
	Writing	19	1,167	13.2	111.6	1,369	12.4	111.3
	Comprehen	39	1,175	30.2	110.5	1,370	30.1	111.2
	Total	83	1,175	65.3	421.1	1,371	64.5	420.8
<b>Grade 5</b>			<b>2006</b>			<b>2007</b>		
<b>C1</b>	Listening	22	328	16.2	104.8	87	14.0	97.4
	Speaking	22	309	16.4	103.0	88	12.0	85.4
	Reading	15	308	11.0	105.1	87	9.1	93.8
	Writing	15	308	10.6	103.5	87	8.1	91.6
	Comprehen	31	328	21.6	102.9	87	18.7	95.6
	Total	74	328	51.9	401.9	88	42.9	381.9
<b>C2</b>	Listening	22	1,066	18.9	114.9	1,210	19.1	115.5
	Speaking	22	1,064	19.8	117.2	1,208	19.9	116.1
	Reading	20	1,067	16.0	116.0	1,215	15.8	116.5
	Writing	19	1,067	14.1	116.7	1,215	13.3	115.8
	Comprehen	39	1,067	32.0	114.6	1,216	32.0	115.1
	Total	83	1,067	68.7	429.6	1,216	67.9	428.9

<b>Grade 6</b>			<b>2006</b>			<b>2007</b>		
<b>Form</b>	<b>Language Domain</b>	<b>RS<sub>Max</sub></b>	<b>N</b>	<b>RS<sub>Mean</sub></b>	<b>SS<sub>Mean</sub></b>	<b>N</b>	<b>RS<sub>Mean</sub></b>	<b>SS<sub>Mean</sub></b>
<b>D1</b>	Listening	22	247	13.6	92.9	71	11.9	87.3
	Speaking	22	239	13.9	91.8	69	9.9	79.9
	Reading	15	245	9.4	91.3	72	7.4	83.7
	Writing	15	245	9.2	91.5	72	7.8	84.3
	Comprehen	32	247	18.9	91.6	72	16.2	85.7
	Total	74	247	45.5	380.8	72	36.4	365.7
<b>D2</b>	Listening	22	1,059	17.8	103.1	1,120	18.0	102.3
	Speaking	22	1,058	19.2	107.3	1,120	18.5	103.8
	Reading	24	1,061	14.5	102.0	1,121	14.3	101.4
	Writing	20	1,062	12.7	102.4	1,119	12.6	101.8
	Comprehen	43	1,065	30.4	101.8	1,123	30.5	100.6
	Total	88	1,065	63.9	404.4	1,124	63.1	402.3
<b>Grade 7</b>			<b>2006</b>			<b>2007</b>		
<b>D1</b>	Listening	22	228	14.1	94.1	82	11.4	86.7
	Speaking	22	220	14.1	93.2	77	10.0	80.7
	Reading	15	229	9.9	93.0	83	8.0	85.6
	Writing	15	228	9.8	94.2	83	8.0	85.7
	Comprehen	32	229	19.9	93.4	83	16.5	86.1
	Total	74	229	47.2	383.9	83	36.6	366.3
<b>D2</b>	Listening	22	945	18.7	106.2	1,092	18.6	104.6
	Speaking	22	940	19.3	107.2	1,090	19.1	106.4
	Reading	24	945	15.8	104.9	1,090	15.5	103.9
	Writing	20	945	13.4	104.6	1,091	13.2	104.0
	Comprehen	44	946	32.7	104.9	1,092	32.3	103.0
	Total	88	946	67.0	409.2	1,093	66.3	407.3
<b>Grade 8</b>			<b>2006</b>			<b>2007</b>		
<b>D1</b>	Listening	22	232	13.8	93.8	78	10.1	84.2
	Speaking	22	225	13.8	92.6	71	7.8	73.8
	Reading	15	230	10.0	93.3	78	7.1	82.4
	Writing	15	229	9.9	94.8	78	7.8	84.3
	Comprehen	32	232	19.8	93.3	78	14.5	83.4
	Total	74	232	46.8	383.6	78	32.1	359.9
<b>D2</b>	Listening	22	851	19.0	107.9	960	18.8	105.5
	Speaking	22	850	19.3	107.6	959	19.1	106.7
	Reading	24	852	16.3	106.4	961	16.0	105.0
	Writing	20	855	13.9	106.4	962	13.5	104.7
	Comprehen	44	855	33.4	106.2	963	32.9	103.9
	Total	88	856	68.2	411.8	963	67.2	408.9

<b>Grade 9</b>			<b>2006</b>			<b>2007</b>		
<b>Form</b>	<b>Language Domain</b>	<b>RS<sub>Max</sub></b>	<b>N</b>	<b>RS<sub>Mean</sub></b>	<b>SS<sub>Mean</sub></b>	<b>N</b>	<b>RS<sub>Mean</sub></b>	<b>SS<sub>Mean</sub></b>
<b>E1</b>	Listening	22	222	12.7	87.8	83	9.9	84.7
	Speaking	22	219	12.0	86.0	80	7.2	74.7
	Reading	15	218	8.4	87.8	84	7.4	81.4
	Writing	15	216	6.5	87.1	84	4.8	80.8
	Comprehen	32	223	17.3	87.0	84	14.1	82.8
	Total	74	224	38.7	376.0	84	28.8	366.0
<b>E2</b>	Listening	22	807	18.6	103.5	903	17.2	101.6
	Speaking	22	798	19.2	105.6	904	17.3	101.6
	Reading	25	807	16.3	102.0	903	15.8	99.7
	Writing	20	798	11.7	102.0	900	11.7	100.2
	Comprehen	44	807	33.1	102.1	908	31.3	99.8
	Total	89	808	65.4	403.3	910	61.5	399.2
<b>Grade 10</b>			<b>2006</b>			<b>2007</b>		
<b>E1</b>	Listening	22	205	13.5	89.4	70	11.2	87.2
	Speaking	22	204	14.0	91.9	70	9.3	81.5
	Reading	15	199	8.7	88.5	70	8.6	85.4
	Writing	15	199	7.1	89.5	70	5.9	85.1
	Comprehen	32	206	18.3	88.2	70	16.5	85.9
	Total	74	207	42.4	380.3	70	35.0	373.0
<b>E2</b>	Listening	22	716	19.2	105.5	782	17.7	103.5
	Speaking	22	713	19.4	106.6	783	17.9	104.5
	Reading	25	716	17.3	104.0	788	16.8	102.0
	Writing	20	719	12.3	103.7	785	12.1	101.8
	Comprehen	45	721	34.4	103.7	790	32.6	101.9
	Total	89	721	67.7	406.1	792	63.9	402.7
<b>Grade 11</b>			<b>2006</b>			<b>2007</b>		
<b>E1</b>	Listening	22	140	14.8	92.5	36	11.7	88.0
	Speaking	22	141	15.1	94.2	36	9.4	80.8
	Reading	15	137	9.7	91.7	36	9.1	87.3
	Writing	15	136	7.8	92.4	36	6.1	85.8
	Comprehen	32	141	20.2	91.5	36	17.5	87.4
	Total	74	142	46.4	385.0	36	36.2	374.3
<b>E2</b>	Listening	22	517	19.1	105.7	644	18.1	104.2
	Speaking	22	506	19.7	108.0	640	18.2	105.3
	Reading	25	517	17.7	105.2	642	17.4	103.4
	Writing	20	516	12.6	104.7	640	12.6	103.0
	Comprehen	45	518	34.9	104.9	647	33.5	103.0
	Total	89	518	68.4	407.9	647	65.6	404.5

Grade 12			2006			2007		
Form	Language Domain	RS <sub>Max</sub>	N	RS <sub>Mean</sub>	SS <sub>Mean</sub>	N	RS <sub>Mean</sub>	SS <sub>Mean</sub>
E1	Listening	22	82	14.4	91.7	20	11.2	87.3
	Speaking	22	82	14.6	92.9	20	10.6	85.2
	Reading	15	80	9.8	92.2	20	9.1	87.7
	Writing	15	80	7.9	92.7	20	6.1	85.8
	Comprehen	32	82	20.3	91.7	20	16.8	86.5
	Total	74	82	46.2	385.0	20	37.0	375.2
E2	Listening	22	431	19.5	107.3	399	17.7	103.1
	Speaking	22	427	19.5	107.5	401	18.4	106.2
	Reading	25	432	17.6	105.3	401	17.4	103.5
	Writing	20	434	12.7	105.0	399	12.6	103.3
	Comprehen	45	435	35.0	105.2	402	33.4	102.8
	Total	89	435	68.6	408.0	408	64.9	403.6

Performance on IELA 2006 and IELA 2007 is summarized in Table 9. This table shows the percent of students in each Total IELA Proficiency category by grade. This table represents students classified as LEP and LEP1 but not LEPX. In addition, this table is not from a matched sample. There are several notable results in this table. The percents of students in each grade combined over the two lowest proficiency categories, Beginning and Advanced Beginning, are fairly stable over grades and over years. The percents in the Intermediate category, although more volatile over grades, are consistent over years. Whereas the sum of the top two categories, Early Fluent and Fluent, is fairly stable, there are some consistent changes over grades in both years. There is a notable decline in the percent Fluent in grades 3, 6, and 9. These are the lowest grade in each of their respective grade clusters, the grades in which students are administered a new form, suggesting a possible "form effect". There are several possible explanations for this effect. First, it could result from the way in which standards were set in the grades that represent transitions between grade clusters (i.e., 2-3, 5-6, 8-9). Second, it could be a result of the fact that, for those students who were tested in 2007 and 2008, there was a significant portion of the test items within a grade cluster that were common from year to year. Across grade clusters, however, there were very few items in common. This familiarity could have made the test more challenging when crossing a grade cluster boundary.

**Table 9. Total IELA Proficiency Level by Grade in 2006 and 2007**

Grade	Percent in each Proficiency Category									
	2006					2007				
	Beg	ABeg	Int	EFI	FI	Beg	ABeg	Int	EFI	FI
<b>K</b>	15.3	12.7	19.2	29.6	23.2	12.9	12.3	21.3	30.0	23.4
<b>1</b>	5.3	8.1	17.8	32.8	36.0	3.0	6.9	18.7	29.6	41.8
<b>2</b>	3.8	3.2	17.1	45.9	30.0	2.3	2.4	15.5	42.8	36.9
<b>3</b>	3.0	5.6	23.9	47.2	20.3	2.8	8.7	25.5	45.6	17.5
<b>4</b>	2.5	5.2	31.9	37.6	22.7	2.9	5.3	30.8	33.4	27.5
<b>5</b>	4.2	5.7	21.7	43.4	24.9	3.5	4.8	21.5	40.7	29.4
<b>6</b>	3.5	4.6	37.0	48.6	6.2	3.7	4.3	35.0	50.0	7.0
<b>7</b>	3.4	3.2	28.3	54.1	10.9	3.3	4.2	24.4	53.3	14.8
<b>8</b>	3.7	5.0	24.2	51.9	15.3	4.3	4.2	21.5	53.6	16.3
<b>9</b>	7.2	6.4	36.1	44.7	5.6	6.0	7.5	36.6	46.6	3.2
<b>10</b>	4.7	5.6	30.5	52.8	6.4	4.8	6.3	32.5	48.4	8.1
<b>11</b>	2.4	6.8	30.6	47.9	12.3	2.8	5.0	29.0	55.5	7.8
<b>12</b>	2.3	3.3	32.5	49.9	12.0	3.3	5.8	26.6	55.8	8.4

## **10. Item Development**

**10.1 Alignment Study.** An alignment study, conducted for the purpose of determining the extent to which the 2006 IELA and 2007 IELA forms were aligned with Idaho English Language Proficiency Standards, resulted in a report, Plan for IELA Item Development. This ‘Plan’ served as the guide for developing additional items. A full copy of the Plan for IELA Item Development/Proposed Test Blueprints is included as an Appendix to this report.

**10.2 Item Development.** Item development took place in April and May 2007. Items were written by experienced item writers contracted by Questar. All item writers had previous experience developing items for an English language learner population. Items were written to specific standards and included both multiple-choice and constructed-response items. All items were edited by two Questar editors, and then entered into the IELA Item Development database. Item development guidelines and supporting sample items were provided to each of the writers as a resource. A copy of the Item Writing Overview and Guidelines is included as an Appendix to this report.

**10.3 Content and Bias Review.** Two IELA New Item Review & Bias/Sensitivity workshops were held (June 12-14, 2007 & August 14-16, 2007) in Boise, Idaho. The details of those reviews will be included in the IELA 2008 Annual Report.

## Appendix A.

### Mountain West Assessment Consortium Foundation Document

#### *Introduction*

The *Mountain West Assessment Consortium Foundation Document* is part of a response to the No Child Left Behind Act (NCLB) of 2001 that mandates assessment of English language learners' progress in attaining proficiency of academic English. Since regular state assessments may not accurately reflect the gains English language learners have made in attaining English proficiency, the Mountain West Assessment Consortium has developed an English language proficiency assessment to serve a dual purpose: to measure students' language proficiency and to measure students' progress toward meeting state standards. Through the development and administration of this assessment, Mountain West Consortium states will satisfy the NCLB requirements for monitoring the development of English proficiency of the English language learners in their public schools.

The *Mountain West Assessment Consortium Foundation Document* describes the elements of language proficiency that are the basis for the Mountain West Assessment Consortium's English Language Proficiency Assessment. The purpose of the assessment is to gauge English language learners' progress in learning to listen to, speak, read, and write in the English language. The assessment follows a developmental progression across and within distinct grade spans. It is based on five communication standards recognized as the linguistic underpinnings of language: phonology, morphology, vocabulary, syntax, and function. The standards have been further detailed in benchmark performance descriptors.

Standards and benchmark descriptors are common elements of any framework that describes what students should know and be able to do. Standards are like umbrellas; they are broad-based, encompassing a set of related skills and/or knowledge bases. Benchmarks are more specific statements that describe discrete tasks students will perform in order to demonstrate knowledge or skills within a standard. For example, under the vocabulary standard in reading, one benchmark descriptor is, "*Reads and understands common idioms.*"

The Mountain West Assessment Consortium English Language Proficiency Assessment includes separate modules for children at these grade spans: kindergarten through early first grade; mid-first grade through second grade; third grade through fifth grade; sixth grade through eighth grade; ninth grade through twelfth grade. Within each of these designated grade spans, assessment items have been developed to evaluate growth in English language acquisition across three broad developmental levels: early acquisition, intermediate, and transitional. The assessment battery modules include test items at each of the three developmental levels across the four modalities of listening, speaking, reading, and writing.

It is important to emphasize the breadth of these developmental levels and to recognize that they are not proficiency levels. The developmental levels of the standards are intentionally broad; they are used simply to make general classifications of test items within the assessment. Proficiency or performance levels specify what a student has achieved or demonstrated *relative to a set of standards*. There may, in fact, be as many as five distinct proficiency levels within these three broad developmental levels. Proficiency or performance levels are determined through standard-setting activities that yield cut-scores within the total range of test scores. There are several ways to determine proficiency levels, and each state that elects to use the Mountain West Assessment Consortium English Language Proficiency Assessment will apply its own process to determine proficiency levels.

Benchmarks have been grouped within five standards to reflect the dimensions of communicative competency:

- Phonology/Orthography standards are used to evaluate students' progress in understanding and correctly manipulating the sound system of English.
- Morphology standards are used to evaluate students' progress in understanding and using the rules of English word formation.
- Vocabulary standards are used to evaluate students' understanding and appropriate use of English words and phrases (semantic knowledge).
- Syntax standards are used to evaluate students' progress in understanding and using the rules of English sentence formation.
- Function/Discourse standards are used to evaluate students' ability to use and comprehend English in various oral and written contexts.

Since elements of some standards must be in place before others develop, the application of these five language standards varies across both grade spans and developmental levels. For example, phonology benchmarks are generally addressed more extensively at the early acquisition level than at intermediate or transitional levels. In addition, the requirements for competency in the four modalities (listening, speaking, reading, and writing) vary so that one modality may emphasize some standards over others. For example, expectations for syntax use are more pronounced in the language production modalities of speaking and writing. Similarly, assessment of function/discourse skills is addressed in greatest depth at the transitional level.

All of the standards and benchmarks included in this document are addressed in the assessment. The majority of the benchmarks are addressed in specific assessment tasks. Other benchmarks are addressed indirectly through holistic acts of listening, speaking, reading, or writing. In the receptive processes of listening or reading, acquisition of some benchmarks is inherent in demonstrations of comprehension of the language presented. Holistic scoring rubrics have been developed to encompass such benchmarks in the language production modalities of speaking and writing.

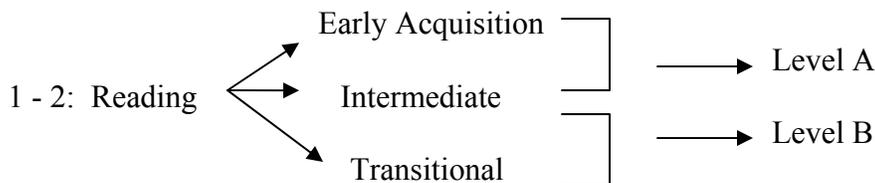
## IELA Item Development

The order in which progress across the four language modalities is assessed also reflects a developmental perspective. The modalities generally considered informal - listening and speaking - precede assessment of the more formal language modalities of reading and writing. Moreover, since a degree of language comprehension generally precedes language production, receptive language skills are addressed before production skills in both informal and formal order in the assessment. Thus, listening skills are assessed first, followed by speaking, reading, and writing skills in that order.

The developmental continuum is also reflected in this assessment in the degree to which language is decontextualized. At the early acquisition level, care has been taken to provide directions that are simple and concrete. Demonstration and practice items are also provided to help students understand what is expected of them. In addition, language in the test directions for intermediate and transitional level items begins to approximate the language found in mainstream assessments.

Spoken English proficiency is assessed in one-to-one settings and all K-1 assessment modules are administered individually. In response to pilot test feedback, all other modules of the Mountain West Assessment Consortium English Language Proficiency Assessment have been designed for group administration. However, the assessment of early acquisition level benchmarks and some intermediate level benchmarks is administrator-led (i.e., the test administrator reads directions and questions to the students). Assessment of intermediate and transitional level benchmarks (Level B; see below) is conducted in typical large-scale assessment format.

The Listening and Speaking assessment modules are designed to be administered in their entirety; each module begins with tasks reflecting early acquisition benchmarks and proceeds through tasks reflecting intermediate and transitional level benchmarks. With the exception of the K-1 measures, Reading and Writing assessment modules are designed so students take either Level A or Level B, but not both. Level A includes assessment tasks that progress from early acquisition benchmarks through early intermediate level. Level B encompasses assessment of intermediate and transitional level benchmarks. An illustration of these relationships appears below. A locator checklist is provided to assist test examiners in determining which test level is appropriate for each student.



**A Quick Guide to the Mountain West Assessment Consortium  
English Proficiency Assessment Foundation Document**

**How the new standards are coded:**

**By modality:**

- Reading
- Speaking
- Writing
- Listening

**By grade span:**

- K-1** Kindergarten through early first grade  
**1-2** Mid-first grade through grade two  
**3-5** Grade three through grade five  
**6-8** Grade six through grade eight  
**9-12** Grade nine through twelve

**By developmental level:**

- Early Acquisition (E)
- Intermediate (I)
- Transitional (T)

**By standard:**

- Phonology/orthography\* (P)
- Morphology (M)
- Vocabulary (V)
- Syntax (S)
- Function/Discourse (F)

**By benchmark:** Benchmarks are the actual performance descriptors. Each is numbered and written as an action statement.

**Examples of how to interpret the coding system:**

- Reading, 3-5 E/P1 (Reading, gr. 3-5, Early Acquisition Level, Phonology benchmark #1): *Distinguishes initial and final sounds in one-syllable words.*
- Speaking, 6-8 I/M1 (Speaking, gr. 6-8, Intermediate Level, Morphology benchmark #1): *Frequently uses appropriate verb and noun endings to indicate possession (e.g., the girl's book), number (e.g., three books), and tense (e.g., she calls; I called).*
- Writing, 9-12 T/S3 (Writing, gr. 9-12, Transitional level, Syntax benchmark #3): *Recognizes dangling participles and revises writing to correct them.*

## IELA Item Development

- Listening, K-1 I/F2 (Listening, gr. K-1, Intermediate Level, Function/Discourse benchmark #2): *Follows two- or three-step directions.*
  - Note: Orthography (spelling) is addressed only

## Appendix B

### Plan for IELA Item Development/Proposed Test Blueprints

*Updated June 6, 2008*

The initial form of the Idaho English Language Assessment (IELA\_2006), first administered in Spring 2006, was based on a test design developed by the Mountain West Assessment Consortium. A second form (IELA\_2007), designed for administration in Spring 2007, was based on the MWAC design but was modified to a degree. The modifications included edits to individual items and the insertion of linking items to equate the two forms.

The purpose of this document is to propose a plan for further revisions to the IELA. The proposed revisions serve both a general purpose, to improve the quality of the assessment overall, and a specific purpose, to improve the alignment of the assessment with Idaho English language development standards. This proposal will be divided into three sections: Design of the current IELA; Considerations that went into the plan for new development; and Proposed revisions for Test Blueprints.

#### **Design of IELA**

The most comprehensive way to characterize a test is in terms of test specifications which typically include several elements: a descriptive component and a test blueprint component. The descriptive element specifies, at minimum, the purpose of the test and the target population. The test blueprint specifies the major content areas covered by the test, the cognitive levels to be assessed, and the numbers of items each form should include within each of these content and cognitive areas. The test blueprint can serve a number of different purposes, but it is essential to provide direction to item writers and other test development professionals. The following paragraphs and associated tables are presented as test specifications for the IELA.

The purpose of the IELA is to assess the English proficiency of limited English proficient students in grades Kindergarten through 12. It was designed to assess English proficiency in four primary language domains, Speaking, Listening, Reading, and Writing. In addition, the results in these individual domains are combined to provide measures of Comprehension (a combination of listening and reading items) and overall English proficiency (a combination of the items from all language domains). Designing a test to assess English proficiency over a broad range of ability from Beginning to Fluent and across a broad range of developmental and academic levels (K-12) is a significant challenge. To accommodate the range of academically and developmentally appropriate content, different forms were designed for each of five grade clusters (K, 1-2, 3-5, 6-8, and 9-12). The range of ability (i.e., English proficiency) within each grade cluster was addressed in the IELA by creating two test levels within each grade cluster (except Kindergarten), one appropriate for first year ELLs and the other appropriate for more advanced ELLs.

The design and content of the 2007 IELA is further elaborated in a series of tables located in the Appendix. The panels of Table A1 show the composition of the forms at each level in terms of

the numbers of each item type in each language domain and the points they represent (in parentheses). The next set of columns in each panel shows the total number of items in each language domain and the number of points they represent. The final columns show the percent of the total test (in terms of the number of points) represented by the items in each language domain. Whereas these tables reveal something about the architecture of the test, they do not provide information, beyond the language domain covered, on the content of the test. One consistent feature of the IELA is that, at each grade cluster in which there are two forms (i.e., all except K), Listening and Speaking items constitute a larger part of the Level 1 test (when compared to Level 2) and Reading and Writing items constitute a larger part of the Level 2 test.

The next table further specifies test content by showing, for each language domain, the alignment to relevant English language development standards and benchmarks. Table A2 shows for each form the number of items (and points they represent) in each language domain by the standard addressed. It is clear from the information presented in Table A2 that the 2006 and 2007 IELA test forms do not address the standards comprehensively. This topic will be covered more fully after consideration of the results of the alignment study.

### **Considerations in Proposing Revisions**

The original design of the MWAC assessment was based on a set of Standards and Objectives as described in the *Mountain West Assessment Foundation Document*. Shortly after implementing the IELA, the Idaho State Board of Education adopted a new set of English language development standards, *Idaho Map of Standards for English Learners: K-12 English Language Development Standards* (henceforth, revised ELD standards). An alignment study was undertaken in order to determine the extent to which the IELA was aligned with the revised ELD standards. As part of the alignment study, a group of reviewers was instructed to focus on the alignment between the revised ELD standards and the 2006 and 2007 versions of the IELA. That alignment was evaluated in terms of the following criteria:

- **Categorical concurrence (CC)** - the extent to which the same or consistent categories of content appear in both the standards and the assessment.
- **Linguistic difficulty level (LDL)** - the extent to which items are written at the linguistic difficulty level of the standard (at least 50% is the acceptability criterion).
- **Range-of-knowledge (ROK)** - the extent to which the span of knowledge expected of students by a standard is the same as, or corresponds to, the span of knowledge students need in order to correctly answer assessment items/activities. This criterion considers the number of objectives within the standard with at least one related assessment item/activity.
- **Balance of representation (BOR)** - indicates the degree to which one objective is given more emphasis on the assessment than another.

Panelists' evaluations of each item were summarized in terms of the extent to which the test forms met these criteria with separate summaries for each language domain (L, S, R, & W) and each grade cluster. For those grade clusters in which multiple forms were administered (1-2, 3-5, 6-8, and 9-12) each form (e.g., C1 and C2) was evaluated separately.

All forms across grade clusters and language domains met the CC and LDL criteria. The ROK criterion was not met for Speaking, Reading, and Writing in any of the forms across grade

## IELA Item Development

clusters and was weakly met for Listening in grades K-2, but not in the other grade clusters. Thus, it would be fair to say that 2006 IELA and 2007 IELA test forms did not represent the range of knowledge represented by the revised ELD standards. Across all grade clusters and forms, the BOR criterion was met in Reading and Writing and met or weakly met in Listening and Speaking. Given that the ROK was below criterion, however, the fact that the BOR criterion was met means only that for those standards/objectives that were addressed by the test, the representation was balanced. The overall finding of the alignment study, then, is that across language domains and grade clusters, the ROK of the test needs to be expanded. This conclusion is corroborated by the information presented in Table A2 which shows the standards that are under or overrepresented by test content. In addition to addressing the ROK, the results of the alignment study indicate that, in order to preserve the balance of representation, newly created items should be balanced across the standards/objectives addressed.

In addition to the formal results produced by the alignment study, the reviewers provided a fairly extensive set of comments, usually on specific items. Those comments have also been taken into account in making recommendations for revising the forms.

Along with the alignment study, the results of the 2006 IELA administration informed the proposal to revise the test. Overall test results, as summarized in the technical report, suggest that the difficulty level of the 2006 IELA was not well matched to the ability of the students assessed in some language domains and grade clusters. Evidence for this assertion can be found in the tables of raw score means by grade level (Table 2, IELA Technical Report, 2006). That table shows, for each grade, test form and language domain the mean raw score obtained and the maximum points available. When the mean raw score is too close to the maximum points, the test may not be providing an accurate measure of student ability. Table 1 summarizes information from the tables of raw score means in the 2006 IELA Technical Report. This table shows the number of instances in which the raw score mean is greater than or equal to 75% of the maximum points.

The numerator in each cell represents the number of instances in which the raw score mean meets or exceeds this criterion. The denominator represents the number of cases in the category. For example, in the 3-5 grade cluster, there were 0 out of 12 instances on the level 1 form (C1) in which the mean raw score exceeded 75% of the available points. By contrast, there were 7 of 12 instances on the level 2 form (C2). The 12 instances for each form originate from the number of grades in a cluster (3 in this case) by 4 language domains. Whereas columns 2 and 3 break the data in a grade cluster down by form, i.e., level 1 vs. 2, columns 4 through 7 break the data down by language domain.

**Table 1. Number of cases where raw score mean exceeds 75% of possible points.**

Grade Cluster	Level 1	Level 2	Listen	Speak	Read	Write	
<b>K</b>	0/4		0/1	0/1	0/1	0/1	0/4
<b>1-2</b>	3/8	5/8	3/4	1/4	3/4	1/4	8/16
<b>3-5</b>	0/12	7/12	3/6	3/6	1/6	0/6	7/24
<b>6-8</b>	0/12	6/12	3/6	3/6	0/6	0/6	6/24
<b>9-12</b>	0/16	8/16	4/8	4/8	0/8	0/8	8/32
<b>Totals</b>	3/52	26/52	13/25	11/25	4/25	1/25	

There are several generalizations that can be made on the basis of data presented in Table 1. First, there appears to be a mismatch between ability and test difficulty at both levels and across the language domains in the Grade 1-2 cluster. Second, looking at the totals in the bottom row of Table 1, it appears that the great majority of cases where the average raw score on a form is greater than or equal to 75% of the maximum points is on the Level 2 forms. Finally, looking again at the bottom row, the problem is much more acute in Listening and Speaking than in Reading and Writing. Because the Listening and Speaking subtests are identical on Level 1 and 2 forms within each grade cluster, it makes sense that those students who are taking the Level 2 form, because of their generally higher level of English proficiency, would score higher on those portions of the test.

These data suggest, then, that in addition to expanding the range of knowledge represented on IELA test forms, the difficulty level of the assessment needs to be adjusted to more closely match the abilities of the students for whom it is intended. In situations where the difficulty of a test is not well matched to the ability of the students being assessed, the problem could be limited to a relatively small set of items or it could be more pervasive. It is useful, therefore, to examine item-level data in those areas where Table 1 indicated there were potential problems. Item level data were examined for both forms administered in the Grade 1-2 cluster and for Listening and Speaking subtests in all grade clusters except Kindergarten. This evaluation revealed that, across the grade clusters, there was a significant percentage of Listening and Speaking items (ranging from 14-80%) with very high p-values ( $p \geq 0.90$ ). And, in addition, approximately 26% of the Reading items on the Grade 1-2 forms had similarly high p-values. In the course of revising the test, these items should be reviewed and either replaced or revised.

**Proposed Revisions**

We propose a set of revisions to the existing IELA forms that will improve alignment with Idaho ELD standards and produce a test that is more appropriate to the ability of the students being tested. The revisions proposed for forms in each grade cluster are shown in Table A3 (Proposed 2008 IELA Test Form Design by Alignment with ELD Standards). The proposed revisions are shown in a format that is similar to Table A2 in order to make clear the consequences for the alignment of the revised forms. Where new items are proposed, a brief description of the items is provided.

Most of the changes proposed have been made to address issues of alignment. In comparing the design of the current test forms, presented in Table A2, to the proposed forms, it is clear that the

## IELA Item Development

proposed test provides better coverage of the standards, thus addressing concerns about the Range of Knowledge raised in the alignment study. In addition, we have attempted to maintain and to improve, where possible, the Balance of Representation. One specific change that is worthy of mention is the addition of a reading fluency task. This task was included because reading fluency is an important indicator of developing reading ability and because it is a part of the revised Idaho ELD standards. It is worthy of mention because it must be individually administered. We propose that it be administered at the same time as the speaking portion of the revised assessment.

Although it is clear from a comparison of Table A2 with Table A3 that there is better coverage of the standards in the proposed design, there remain some cells with no items. This situation is due, in part, to the structure of the Idaho ELD standards which include:

- Four ELD Standards, one each for Listening, Speaking, Reading, and Writing;
- Seven goals, one each for Listening and Speaking, two for Reading, and three for Writing; and,
- A variable number of Objectives under each Goal.

The three numbers preceding the descriptors in the first line of each panel (e.g., 3.1.3) correspond to the standard, goal, and objective, respectively. The proposed revisions to the IELA cover all four standards and 6 of the 7 goals. The goal which is not specifically covered is 4.1 Writing Process. This goal is concerned with the ability to use planning strategies to write, edit, and revise a draft. For example, one objective is to “fill in prewriting organizer, write a draft, make basic revisions, and edit the draft for some writing conventions.” Testing this entire process would be time consuming and rating students’ writing processes with consistency (as opposed to rating the final product) would be difficult. This goal is best assessed in the classroom environment. The primary reason why some cells in Table A3 remain empty, however, is that we have not endeavored to test all objectives. The decision about whether or not to test an objective was based on several considerations: overall test length and the academic appropriateness of the objective. An attempt was made to test as many objectives as possible without significant increases in the overall length of the assessment. In several cases, it was deemed more academically appropriate to assess an objective at one level of a grade cluster but not at the other. For example, phonological awareness (standard 3.1.3) is assessed in B1 but not B2 and Reading with fluency (standard 3.1.7) is assessed in B2 but not B1.

Another set of proposed changes is intended to better align the difficulty of the test with the ability of students being assessed. It is more difficult to evaluate, in advance, the likely impact of these changes, but data on existing items can be used to inform revisions to targeted items and/or the development of new items. A more general issue concerning changes to the Listening and Speaking subtests will be addressed in a subsequent section.

Proposed revisions are shown in terms of changes to the 2007 IELA rather than the 2006 IELA because, overall, the 2007 IELA is a better test than 2006 IELA. In building the 2007 IELA forms, the items were chosen more selectively and some modifications were made to improve item format. For example, the Listening passages are being read only once rather than twice as they were in the 2006 IELA. Because the structure of the 2006 IELA and the 2007 IELA were similar, however, the same changes to 2006 IELA or to some combination of the 2006 and 2007 forms would similarly improve alignment to standards.

**Plan to implement the proposed revisions.**

- New item development. Proposed items will be developed and presented to committees in Idaho for content and bias review. The numbers of items to be developed will be addressed below.
- Field testing. The necessity for a field test of new items has been discussed by ID and Questar (TASA) and it has been resolved that new items must be field tested prior to being eligible as operational items. The proper format for a field test has also been discussed and it has been established that a stand-alone field test is not feasible. New items, therefore, following a content and bias/sensitivity review, will be embedded into operational forms and field tested therein. Thus forms administered in 2008 will have an operational portion and a field test portion. Once the new items have been field tested, those that have acceptable item parameters will become part of the item pool from which future IELA forms will be built.
- 2008 IELA Operational Forms. Although the initial proposal was to modify IELA 2007 forms, the proposal has been revised to build operational forms using items from the 2006 and 2007 IELA forms. This strategy has several desirable features. First, there will be no exact replicas of prior forms, i.e., neither the 2006 or 2007 form will be duplicated in full. Second, it presents the opportunity to standardize the format of items not previously standardized, e.g., some of the listening items first used in 2006 have not yet been changed to the new format in which passages are read only once. Third, this strategy provides the opportunity to select from the available item pool those items that are most appropriate to the ability of students being assessed. The operational portion of the forms will be no longer than the 2006 and 2007 forms and, in many instances, it will be shorter. The number of field test items will be small relative to the size of the language domain subtests. Field test items will not be segregated or identified in any way. They will be inserted into the forms adjacent to similar items and in locations that do not interfere with the flow of the assessment.

The proposed design for the 2008 forms is shown in Table 2. The cells in this table show the number of points in the operational portion of each subtest plus the number of points represented by field test items. In addition the table shows, for each grade cluster and form, the number of proposed field test forms.

**Table 2. Proposed Field Test Form Design**

<b>Grade Cluster</b>	<b>Form</b>	<b>FT Forms</b>	<b>Listen</b>	<b>Speak</b>	<b>Read</b>	<b>Write</b>	<b>Total</b>
<b>K</b>	<b>A</b>	<b>2</b>	<b>20+4</b>	<b>20+5</b>	<b>26+8</b>	<b>22*</b>	<b>88+17</b>
<b>1-2</b>	<b>B1</b>	<b>1</b>	<b>20+4</b>	<b>20+6</b>	<b>15+8</b>	<b>15+6</b>	<b>70+24</b>
	<b>B2</b>	<b>3</b>	<b>20+4</b>	<b>20+6</b>	<b>20+8</b>	<b>20+4</b>	<b>80+22</b>
<b>3-5</b>	<b>C1</b>	<b>1</b>	<b>20+4</b>	<b>20+4</b>	<b>15+6</b>	<b>15+6</b>	<b>70+20</b>
	<b>C2</b>	<b>3</b>	<b>20+4</b>	<b>20+4</b>	<b>20+6</b>	<b>20+6</b>	<b>80+20</b>
<b>6-8</b>	<b>D1</b>	<b>1</b>	<b>20+6</b>	<b>20+4</b>	<b>15+6</b>	<b>15+6</b>	<b>70+22</b>
	<b>D2</b>	<b>3</b>	<b>20+6</b>	<b>20+4</b>	<b>24+6</b>	<b>20+6</b>	<b>84+22</b>
<b>9-12</b>	<b>E1</b>	<b>1</b>	<b>20+4</b>	<b>20+4</b>	<b>15+6</b>	<b>15+6</b>	<b>70+20</b>
	<b>E2</b>	<b>3</b>	<b>20+4</b>	<b>20+4</b>	<b>24+6</b>	<b>20+5</b>	<b>84+19</b>

Alignment of the 2008 forms will not necessarily be an improvement over 2006 or 2007 since none of the newly developed items, developed for the purpose of improving alignment, will be operational. When the newly developed items are incorporated into operational forms in 2009 and 2010, however, the alignment of the IELA forms with ELD standards will be improved and comparable to that shown in Table A4 of the proposal.

- Number of Field Test Forms. One issue that must be decided is whether to build one or multiple field test forms for each level in each grade cluster (i.e., one C1 and one C2). Because of the potentially small numbers of students being administered the level 1 form in each grade cluster, it is really only feasible to consider whether to build multiple level 2 forms in each grade cluster. There are advantages and disadvantages to each approach. One field test form per level per grade cluster simplifies test administration since only one set of instructions needs to be developed for each test level. The disadvantage of this approach is that only a small set of new items can be field tested within a given year. Creating multiple level 2 field test forms per grade cluster would complicate test administration but would yield a higher number of items in a given period of time. (It was decided that six (6) level 2 field test forms per grade cluster would be developed.)
- Number of Items to Develop. The proposal is to develop approximately four times the number of items proposed in Table A3 in order to: 1) allow for items that do not survive item review; 2) develop some new items for all subtests and levels even those in which the analysis in Table A3 did not indicate a need for items to improve alignment; 3) develop additional Listening and Speaking items to accommodate the change proposed in the following paragraph.

## IELA Item Development

- **Listening and Speaking Subtests.** It is not clear that the changes to Listening and Speaking subtests proposed in Table A3 will be sufficient to address the mismatch between ability and test difficulty that is apparent on level 2 forms. It is proposed, therefore, that beginning in 2008, the Listening and Speaking portions of the different test levels (i.e., 1 and 2) in each grade cluster not be identical. There will be a significant amount of overlap, but there will also be items that are not common to the two levels. In this way, items that are more appropriate to the abilities of the students being assessed can be incorporated into the test at each level. Concomitant with this change, the length of the operational portion of both the Listening and Speaking tests has been reduced from 22 to 20 points.
- **Timetable.** Item development would begin as soon as possible and item/bias review would take place in early June 2007. Under that schedule, we would have enough time to incorporate items into the 2008 IELA forms.
- **Future Forms.** From the items that are available from the existing pool as well as items that are field tested in 2008, two operational forms, to be administered in 2009 and 2010, could be constructed. Those forms will be similar in length to the operational portion of the proposed field test forms.

### **Content and Bias/Sensitivity Review Panels**

All new items will be reviewed for potential bias and fairness as well as for content. The State has proposed that these reviews be conducted by the same panel of individuals. Although these reviews are often completed by separate panels, Questar does not feel that this approach will compromise the results. It is recommended, however, that the panel(s) be directed to separate their considerations and judgments about issues of content from those of potential bias. It is also recommended that the State determine, in advance, that this approach will meet federal guidelines. The increased number of items will require additional person days for review. It is estimated that the bias and content review of the newly developed items will require 72 person days (2 panels of 12 for three days each). (The item reviews took place in June 12-14 and August 14-16, 2007.)

### **Budgetary Impact**

Given that the scope of the item development work has changed, there will be changes in the originally proposed item development budget. The original proposal was to develop 286 items. The current proposal is for 555 items. The original proposal for item review included 4 panels of 12 for one day or 48 person days. The current proposal includes 2 panels of 12 for 3 days or 72 person days. (555 items overall were developed.)

**Table A1. IELA\_2007 Test Design by Item Type**

**Form A (Grade K)**

Language Domain	MC	CR1	CR2	CR4	Total Items	Total Points	% of Test
Listening	5 (5)	15 (15)			20	20	21%
Speaking		10 (10)	2 (4)	2 (8)	14	22	23%
Reading	12 (12)	18 (18)			30	30	32%
Writing	checklist					22	23%
						94	

(MC- Multiple choice; CRx- Constructed response with point value=x)

**Form B (Grades 1-2)**

Language Domain	MC	CR1	CR2	CR4	Total Items	Points B1 Test	Points B2 Test	% of B1 Test	% of B2 Test
Listening	22				22	22	22	30%	26%
Speaking		10 (10)	2 (4)	2 (8)	14	22	22	30%	26%
Reading B1	15				15	15	-	20%	
Writing B1		11 (11)	2 (4)		13	15	-	20%	
Reading B2	20				20	-	20		24%
Writing B2		10 (10)	1 (2)	2 (8)	13	-	20		24%
						74	84		

**Form C (Grades 3-5)**

Language Domain	MC	CR1	CR2	CR4	Total Items	Points C1 Test	Points C2 Test	% of C1 Test	% of C2 Test
Listening	22				22	22	22	30%	27%
Speaking		10 (10)	2 (4)	2 (8)	14	22	22	30%	27%
Reading C1	15				15	15		20%	
Writing C1	4	5 (5)	1 (2)	1 (4)	11	15		20%	
Reading C2	18		1 (2)		19		20		24%
Writing C2	9		1 (2)	2 (8)	12		19		23%
						74	83		

**Form D (Grades 6-8)**

Language Domain	MC	CR1	CR2	CR4	Total Items	Points D1 Test	Points D2 Test	% of D1 Test	% of D2 Test
Listening	22				22	22	22	30%	25%
Speaking		10 (10)	2 (4)	2 (8)	14	22	22	30%	25%
Reading D1	15				15	15		20%	
Writing D1	5	4 (4)	1 (2)	1 (4)	11	15		20%	
Reading D2	18		1 (2)	1 (4)	20	-	24		27%
Writing D2	10		1 (2)	2 (8)	13	-	20		23%
						74	88		

IELA Item Development

**Form E (Grades 9-12)**

<b>Language Domain</b>	<b>MC</b>	<b>CR1</b>	<b>CR2</b>	<b>CR4</b>	<b>Total Items</b>	<b>Points E1 Test</b>	<b>Points E2 Test</b>	<b>% of E1 Test</b>	<b>% of E2 Test</b>
Listening	22				22	22	22	30%	25%
Speaking		10 (10)	2 (4)	2 (8)	14	22	22	30%	25%
Reading E1	15				15	15		20%	
Writing E1	7	2 (2)	1 (2)	1 (4)	11	15		20%	
Reading E2	19		1 (2)	1 (4)	21		25		28%
Writing E2	10		1 (2)	2 (8)	13		20		22%
						74	89		

**Table A2. IELA\_2007 Test Design by Alignment with ELD Standards**

**2007 Form A - Listening**

	1.1.1 Follow Oral Directions	1.1.2 Understand Social & Academic Conversations	1.1.3 Understand Key Ideas of Information Presented Orally	NA	Total
<b>Items</b>	3	0	13	4	20 items
<b>Points</b>	3	0	13	4	20 pts

Note: NA = non-aligned items

**2007 Form A - Speaking**

	2.1.1 Ask & answer questions	2.1.2 Communicate information orally	2.1.3 Retell stories or experiences	NA	Total
<b>Items</b>	2	8	2	2	14 items
<b>Points</b>	2	10	8	2	22 pts

**2007 Form A - Reading**

	3.1.1 Text features	3.1.2 Graphic features	3.1.3 Phonologica l awareness	3.1.4 Knowledg e of syllables	3.1.5 Determin e meaning of words	3.1.6 Synonym s, antonyms , homonym s	3.1.7 Read with fluency	3.2 Reading comprehensi on	NA	Total
<b>Items</b>	0	0	18	3	4	0	0	3	2	30
<b>Points</b>	0	0	18	3	4	0	0	3	2	30

**2007 Form A - Writing**

Checklist

IELA Item Development

**2007 Form B - Listening**

	1.1.1 Follow Oral Directions	1.1.2 Understand Social & Academic Conversations	1.1.3 Understand Key Ideas of Information Presented Orally	NA	Total
<b>Items</b>	5	0	15	2	22 items
<b>Points</b>	5	0	15	2	22 pts

**2007 Form B - Speaking**

	2.1.1 Ask & answer questions	2.1.2 Communicate information orally	2.1.3 Retell stories or experiences	NA	Total
<b>Items</b>	0	6	5	3	14 items
<b>Points</b>	0	7	12	3	22 pts

**2007 Form B1 - Reading**

	3.1.1 Text features	3.1.2 Graphic features	3.1.3 Phonologic al awareness	3.1.4 Knowledg e of syllables	3.1.5 Determin e meaning of words	3.1.6 Synonym s, antonyms , homonym s	3.1.7 Read with fluenc y	3.2.1 Reading comp.: Directions	3.2.2 Reading comp.: Expositor y	3.2.3 Readin g comp.: Fiction	NA	Total
<b>Items</b>	0	0	0	0	2	0	0	2		6	5	15
<b>Points</b>	0	0	0	0	2	0	0	2		6	5	15

**2007 Form B2 - Reading**

	3.1.1 Text features	3.1.2 Graphic features	3.1.3 Phonologic al awareness	3.1.4 Knowledg e of syllables	3.1.5 Decode & Determin e meaning of words	3.1.6 Synonym s, antonyms , homonym s	3.1.7 Read with fluenc y	3.2.1 Reading comp.: Directions	3.2.2 Reading comp.: Expositor y	3.2.3 Readin g comp.: Fiction	NA	Total
<b>Items</b>	0	0	0	0	2	0	0	2	6	10	0	20
<b>Points</b>	0	0	0	0	2	0	0	2	6	10	0	20

IELA Item Development

**2007 Form B1 - Writing**

	4.1.1 Writing Process	4.2.1 Write Narrative s	4.2.2 Write Reports	4.3.1 Spell words correctly	4.3.2 Capitalizati on & punctuation	4.3.3 Use grammatic al forms	NA	Total
<b>Items</b>	0	1	1	7	2	2	0	13
<b>Points</b>	0	2	2	7	2	2	0	15

**2007 Form B2 - Writing**

	4.1.1 Writing Process	4.2.1 Write Narrative s	4.2.2 Write Reports	4.3.1 Spell words correctly	4.3.2 Capitalizati on & punctuation	4.3.3 Use grammatic al forms	NA	Total
<b>Items</b>		1	1	3	3	5		13
<b>Points</b>		4	4	3	3	6		20

IELA Item Development

**2007 Form C - Listening**

	1.1.1 Follow oral directions	1.1.2 Understand social & academic conversations	1.1.3 Understand main idea of information presented orally	NA	Total
<b>Items</b>	2	0	13	5	22
<b>Points</b>	2	0	13	5	22

**2007 Form C - Speaking**

	2.1.1 Ask & answer questions	2.1.2 Communicate information orally	2.1.3 Plan oral presentations	2.1.4 Deliver oral presentations	NA	Total
<b>Items</b>	1*	9	0	2	2	14
<b>Points</b>	1	11	0	8	2	22

**2007 Form C1 - Reading**

	3.1.1 Text features	3.1.2 Graphic features	3.1.3 Phonological awareness	3.1.4 Knowledge of syllables	3.1.5 Decode & determine meaning of words	3.1.6 Synonyms, antonyms, homonyms	3.1.7 Read with fluency	3.2.1 Reading comp.: Directions	3.2.2 Reading comp.: Main idea (expository)	3.2.3 Reading comp.: draw conclusions	3.2.4 Reading comp.: Fiction (plot, etc.)	NA	Total
<b>Items</b>	1	0	2	0	1	1	0	2	3	2	3	0	15
<b>Points</b>	1	0	2	0	1	1	0	2	3	2	3	0	15

**2007 Form C2 - Reading**

	3.1.1 Text features	3.1.2 Graphic features	3.1.3 Phonological awareness	3.1.4 Knowledge of syllables	3.1.5 Decode & determine meaning of words	3.1.6 Synonyms, antonyms, homonyms	3.1.7 Read with fluency	3.2.1 Reading comp.: Directions	3.2.2 Reading comp.: Main idea (expository)	3.2.3 Reading comp.: draw conclusions	3.2.4 Reading comp.: Fiction (plot, etc.)	NA	Total
<b>Items</b>	1	0	0	0	0	2	0	3	4	4	5		19

IELA Item Development

<b>Points</b>	1	0	0	0	0	2	0	3	5	4	5		20
---------------	---	---	---	---	---	---	---	---	---	---	---	--	----

**2007 Form C1 - Writing**

	4.1.1 Writing Process	4.2.1 Write Narrative s	4.2.2 Write Reports	4.3.1 Spell words correctly	4.3.2 Write a variety of sentence types	4.3.3 Capitalizati on & punctuation	4.3.4 Use grammatic al forms	NA	Total
<b>Items</b>	0	2	0	6	0	1	2		11
<b>Points</b>	0	6	0	6	0	1	2		15

**2007 Form C2 - Writing**

	4.1.1 Writing Process	4.2.1 Write Narrative s	4.2.2 Write Reports	4.3.1 Spell words correctly	4.3.2 Write a variety of sentence types	4.3.3 Capitalizati on & punctuation	4.3.4 Use grammatic al forms	NA	Total
<b>Items</b>	0	1	2	3	0	1	5		12
<b>Points</b>	0	2	8	3	0	1	5		19

IELA Item Development

**2007 Form D - Listening**

	1.1.1 Follow oral directions	1.1.2 Understand social & academic conversations	1.1.3 Understand main idea of information presented orally	NA	Total
<b>Items</b>	3		15	4	22
<b>Points</b>	3		15	4	22

**2007 Form D - Speaking**

	2.1.1 Ask & answer questions	2.1.2 Communicate information orally	2.1.3 Organize oral presentations	2.1.4 Deliver oral presentations	NA	Total
<b>Items</b>	4	8	0	2		14
<b>Points</b>	4	10	0	8		22

**2007 Form D1 - Reading**

	3.1.1 Text features	3.1.2 Graphic features	3.1.3 Phonological awareness	3.1.4 Knowledge of syllables	3.1.5 Decode & determine meaning of words	3.1.6 Synonyms, antonyms Multiple meanings	3.1.7 Read with fluency	3.2.1 Reading comp.: Directions	3.2.2 Reading comp.: Main idea (expository)	3.2.3 Reading comp.: draw conclusions	3.2.4 Reading comp.: Fiction (plot, etc.)	NA	Total
<b>Items</b>	0	2	2	0	2	0	0	2	3	1	2	1	15
<b>Points</b>	0	2	2	0	2	0	0	2	3	1	2	1	15

**2007 Form D2 - Reading**

	3.1.1 Text features	3.1.2 Graphic features	3.1.3 Phonological awareness	3.1.4 Knowledge of syllables	3.1.5 Decode & determine meaning of words	3.1.6 Synonyms, antonyms Multiple meanings	3.1.7 Read with fluency	3.2.1 Reading comp.: Directions	3.2.2 Reading comp.: Main idea (expository)	3.2.3 Reading comp.: draw conclusions	3.2.4 Reading comp.: Fiction (plot, etc.)	NA	Total
<b>Items</b>	1	0	0	0	2	2	0	2	5	4	3	1	20
<b>Points</b>	1	0	0	0	2	2	0	2	8	5	3	1	24

IELA Item Development

**2007 Form D1 - Writing**

	4.1.1 Writing Process	4.2.1 Write Narrative s	4.2.2 Write Reports	4.3.1 Spell words correctly	4.3.2 Write a variety of sentence types	4.3.3 Capitalizati on & punctuation	4.3.4 Use grammatic al forms	NA	Total
<b>Items</b>	0	1	1	5	0	2	0	2	11
<b>Points</b>		2	4	5	0	2	0	2	15

**2007 Form D2 - Writing**

	4.1.1 Writing Process	4.2.1 Write Narrative s	4.2.2 Write Reports	4.3.1 Spell words correctly	4.3.2 Write a variety of sentence types	4.3.3 Capitalizati on & punctuation	4.3.4 Use grammatic al forms	NA	Total
<b>Items</b>	5*	1	1	4	0	4	1	1	13
<b>Points</b>	5*	4	4	4	0	4	1	2	20

\*4 of these items are also linked to a second standard, therefore they appear twice in the table. They were not, however, counted twice in the Totals.

IELA Item Development

**2007 Form E - Listening**

	1.1.1 Follow oral directions	1.1.2 Understand social & academic conversations	1.1.3 Understand main idea of information presented orally	NA	Total
<b>Items</b>	3		13	6	22
<b>Points</b>	3		13	6	22

**2007 Form E - Speaking**

	2.1.1 Ask & answer questions	2.1.2 Communicate information orally	2.1.3 Organize oral presentations	2.1.4 Deliver oral presentations	NA	Total
<b>Items</b>	2	8	0	2	2	14
<b>Points</b>	2	10	0	8	2	22

**2007 Form E1 - Reading**

	3.1 Reading Process	3.1.1 Text features	3.1.2 Graphic features	3.1.3 Phonological awareness	3.1.4 Decode words using word parts	3.1.5 Use context to determine meaning of words	3.2.1 Reading comp.: Directions	3.2.2 Reading comp.: Main idea (expository)	3.2.3 Reading comp.: draw conclusions	3.2.4 Reading comp.: Fiction (plot, etc.)	NA	Total
<b>Items</b>	1	0	2	1	0	2	2*	3	0	3	2	15
<b>Points</b>	1	0	2	1	0	2	2*	3	0	3	2	15

\*1 item is already counted as a graphic features item. Therefore, do not count twice for Totals.

**2007 Form E2 - Reading**

	3.1 Reading Process	3.1.1 Text features	3.1.2 Graphic features	3.1.3 Phonological awareness	3.1.4 Decode words using word parts	3.1.5 Use context to determine meaning of words	3.2.1 Reading comp.: Directions	3.2.2 Reading comp.: Main idea (expository)	3.2.3 Reading comp.: draw conclusions	3.2.4 Reading comp.: Fiction (plot, etc.)	NA	Total

IELA Item Development

<b>Items</b>		1	0	0	0	7	2	5	3	1	1	21
<b>Points</b>		1	0	0	0	7	2	5	7	1	1	25

**2007 Form E1 - Writing**

	4.1.1 Writing Process	4.2.1 Write Narrative s	4.2.2 Write Reports	4.3.1 Spell words correctly	4.3.2 Write a variety of sentence types	4.3.3 Capitalizati on & punctuation	4.3.4 Use grammatic al forms	NA	Total
<b>Items</b>	0	0	1	4	1	2	3	0	11
<b>Points</b>	0	0	4	4	2	2	3	0	15

**2007 Form E2 - Writing**

	4.1.1 Writing Process	4.2.1 Write Narrative s	4.2.2 Write Reports	4.3.1 Spell words correctly	4.3.2 Write a variety of sentence types	4.3.3 Capitalizati on & punctuation	4.3.4 Use grammatic al forms	NA	Total
<b>Items</b>	5*	0	1	4	1	2	4	1**	13
<b>Points</b>	5*	0	4	4	2	2	4	4	20

\*These are not counted in the Totals because the same items are also counted in 4.3.1 & 4.3.4.

\*\*Persuasive essay.

**Table A3. Proposed IELA\_2008 Test Form Design by Alignment with ELD Standards**

**2008 Form A - Listening (proposed)**

	1.1.1 Follow oral directions	1.1.2 Understand social & academic conversations	1.1.3 Understand key ideas of information presented orally	NA	Total
<b>Items</b>	3	4	13	0	20
<b>Points</b>	3	4	13	0	20
<b>New Items</b>		4			4

New items:

- 2 CR1 Understand Social Conversations
- 2 CR1 Understand Academic Conversations

**2008 Form A - Speaking (proposed)**

	2.1.1 Ask & answer questions	2.1.2 Communicate information orally	2.1.3 Retell stories or experiences	NA	Total
<b>Items</b>	5	7	2		14
<b>Points</b>	5	9	8		22
<b>New Items</b>	3	2			5

New items:

- 2 CR1 Ask Questions
- 1 CR1 Answer Questions (same type as item #3 but higher level)
- 1 CR1 Vocab Question (same type as #6 but higher level vocab)
- 1 CR2 Retell Shared Information - to replace item #11

IELA Item Development

**2008 Form A - Reading (proposed)**

	3.1.1 Text feature s	3.1.2 Graphic features	3.1.3 Phonologica l awareness	3.1.4 Knowledg e of syllables	3.1.5 Determin e meaning of words	3.1.6 Synonym s, antonyms , homonym s	3.1.7 Read with fluency	3.2 Reading comprehensi on	NA	Total
<b>Items</b>	2	0	12	3	4	2	4	3		30
<b>Points</b>	2	0	12	3	4	2	4	3		30
<b>New Items</b>	2					2	4			8

New items:

- 1 CR1 Text Features [circle first word in sentence]
- 1 MC Text Features [which picture shows the front cover of the book]
- 2 MC Synonyms/Antonyms (1 synonym, 1 antonym)
- 4 CR1 Read Words Aloud

IELA Item Development

**2008 Form B - Listening (proposed)**

	1.1.1 Follow oral directions	1.1.2 Understand social & academic conversations	1.1.3 Understand key ideas of information presented orally	NA	Total
<b>Items</b>	5	6	11		22
<b>Points</b>	5	6	11		22
<b>New Items</b>		6			6

New items:

- 2 MC Understand Social Conversations
- 4 MC Understand Academic Conversations

Edit:

- Shorten Listening passages. [Both reviewers and examiners gave feedback that passages are too long for this grade level.]

**2008 Form B - Speaking (proposed)**

	2.1.1 Ask & answer questions	2.1.2 Communicate information orally	2.1.3 Retell stories or experiences	NA	Total
<b>Items</b>	6	6	4		16
<b>Points</b>	6	7	12		25
<b>New Items</b>	6	2			8

New items:

- 3 CR1 Ask Questions
- 3 CR1 Answer Questions (same type as item Form A #3)
- 2 CR1 Vocab Questions (higher level vocab - nouns)

Edit:

- Replace 2 Vocab (2.1.2) items with higher level vocab
- Drop #4 or #5 (retell details from story)

IELA Item Development

**2008 Form B1 - Reading (proposed)**

	3.1.1 Text features	3.1.2 Graphic features	3.1.3 Phonologic al awareness	3.1.4 Knowledg e of syllables	3.1.5 Determin e meaning of words	3.1.6 Synonym s, antonyms , homonym s	3.1.7 Read with fluenc y	3.2.1 Reading comp.: Directions	3.2.2 Reading comp.: Expositor y	3.2.3 Readin g comp.: Fiction	NA	Total
<b>Items</b>	1	0	3	0	4	2	0	2	0	6	0	18
<b>Points</b>	1	0	3	0	4	2	0	2	0	6	0	18
<b>New Items</b>	1		3		2	2						8

New items:

- 1 MC Text Features [book title]
- 2 MC Synonyms/Antonyms (1 synonym, 1 antonym)
- 3 MC Match letter sound to printed letter
- 2 MC Match the word to the picture (Beginner Level words)

**2008 Form B2 - Reading (proposed)**

	3.1.1 Text features	3.1.2 Graphic features	3.1.3 Phonologic al awareness	3.1.4 Knowledg e of syllables	3.1.5 Decode & Determin e meaning of words	3.1.6 Synonym s, antonyms , homonym s	3.1.7 Read with fluenc y	3.2.1 Reading comp.: Directions	3.2.2 Reading comp.: Expositor y	3.2.3 Readin g comp.: Fiction	NA	Total
<b>Items</b>	1	1	0	0	2	2	1	2	6	10		25
<b>Points</b>	1	1	0	0	2	2	6	2	6	10		30
<b>New Items</b>	1	1				2	1					5

New items:

- 1 MC Text Features [author]
- 1 MC Bar Graph
- 2 MC Synonyms/Antonyms (1 synonym, 1 antonym)
- 1 ER Fluency Passage (read aloud for 1 min.; score on cwpm)\*

\*The Fluency task would be administered during the individually administered Speaking Test.

IELA Item Development

**2008 Form B1 - Writing (proposed)**

	4.1.1 Writing Process	4.2.1 Write Narrative s	4.2.2 Write Reports	4.3.1 Spell words correctly	4.3.2 Capitalizati on & punctuation	4.3.3 Use grammatic al forms	NA	Total
<b>Items</b>	0	1	1	7	2	3	0	14
<b>Points</b>	0	2	2	7	2	3	0	16
<b>New Items</b>						1		1

New items:

- 1 CR1 Cloze Item [-ing verb]

**2008 Form B2 - Writing (proposed)**

	4.1.1 Writing Process	4.2.1 Write Narrative s	4.2.2 Write Reports	4.3.1 Spell words correctly	4.3.2 Capitalizati on & punctuation	4.3.3 Use grammatic al forms	NA	Total
<b>Items</b>		1	1	3	3	5		13
<b>Points</b>		4	4	3	3	6		20
<b>New Items</b>								0

IELA Item Development

**2008 Form C - Listening (proposed)**

	1.1.1 Follow oral directions	1.1.2 Understand social & academic conversations	1.1.3 Understand key ideas of information presented orally	NA	Total
<b>Items</b>	2	6	13		23
<b>Points</b>	2	6	13		23
<b>New Items</b>		6			6

New items:

- 2 MC Understand Social Conversations
- 4 MC Understand Academic Conversations

**2008 Form C - Speaking (proposed)**

	2.1.1 Ask & answer questions	2.1.2 Communicate information orally	2.1.3 Plan oral presentations	2.1.4 Deliver oral presentations	NA	Total
<b>Items</b>	6	10	0	2	0	18
<b>Points</b>	6	12	0	8	0	26
<b>New Items</b>	6	1				7

New items:

- 3 CR1 Ask Questions
- 3 CR1 Answer Questions (same type as item Form A #3)
- 1 CR1 Academic-related vocab - similar in difficulty to item #9

Replace:

- \*Item #6 (answer questions) is too easy. Replace with more challenging item of same type (i.e., answer questions)

IELA Item Development

**2008 Form C1 - Reading (proposed)**

	3.1.1 Text featur es	3.1.2 Graphi c featur s	3.1.3 Phonologic al awareness	3.1.4 Knowledg e of syllables	3.1.5 Decode & determin e meaning of words	3.1.6 Synonym s, antonyms , homonym s	3.1.7 Read with fluenc y	3.2.1 Reading comp.: Direction s	3.2.2 Reading comp.: Main idea (expositor y)	3.2.3 Reading comp.: draw conclusio ns	3.2.4 Readin g comp.: Fiction (plot, etc.)	NA	Total
<b>Items</b>	1	1	2	0	1	1	1	2	3	2	3	0	17
<b>Points</b>	1	1	2	0	1	1	6	2	3	2	3	0	22
<b>New Items</b>		1					1						2

New items:

- 1 MC Graphics Features (e.g., understand diagram, map, chart)
- 1 CR6 Fluency

**2008 Form C2 - Reading (proposed)**

	3.1.1 Text featur s	3.1.2 Graphi c featur s	3.1.3 Phonologic al awareness	3.1.4 Knowledg e of syllables	3.1.5 Decode & determin e meaning of words	3.1.6 Synonym s, antonyms , homonym s	3.1.7 Read with fluenc y	3.2.1 Reading comp.: Direction s	3.2.2 Reading comp.: Main idea (expositor y)	3.2.3 Reading comp.: draw conclusio ns	3.2.4 Readin g comp.: Fiction (plot, etc.)	NA	Total
<b>Items</b>	1	1	0	0	0	2	1	3	4	4	5		21
<b>Points</b>	1	1	0	0	0	2	6	3	5	4	5		27
<b>New Items</b>		1					1						2

New items:

- 1 MC Graphics Features (e.g., understand diagram, map, chart)
- 1 CR6 Fluency passage

IELA Item Development

**2008 Form C1 - Writing (proposed)**

	4.1.1 Writing Process	4.2.1 Write Narrative s	4.2.2 Write Reports	4.3.1 Spell words correctly	4.3.2 Write a variety of sentence types	4.3.3 Capitalizati on & punctuation	4.3.4 Use grammatic al forms	NA	Total
<b>Items</b>	0	2	0	6	2	1	4		15
<b>Points</b>	0	6	0	6	4	1	4		21
<b>New Items</b>					2		2		4

New items:

- 1 MC Grammar: Identify noun/verb in sentence
- 1 MC Grammar: Choose correct verb form
- 1 CR2 Write a simple declarative sentence (in response to picture prompt)
- 1 CR2 Write a simple interrogative sentence beginning with a given word (e.g., Where...)

**2008 Form C2 - Writing (proposed)**

	4.1.1 Writing Process	4.2.1 Write Narrative s	4.2.2 Write Reports	4.3.1 Spell words correctly	4.3.2 Write a variety of sentence types	4.3.3 Capitalizati on & punctuation	4.3.4 Use grammatic al forms	NA	Total
<b>Items</b>	0	1	2	3	2	1	6		15
<b>Points</b>	0	2	8	3	3	1	6		23
<b>New Items</b>					2		1		3

New items:

- 1 CR2 Write an interrogative sentence
- 1 MC identify a declarative sentence [OR 1 MC identify adjective/pronoun in a sentence]
- 1 MC identify noun/verb in a sentence

IELA Item Development

**2008 Form D - Listening (proposed)**

	1.1.1 Follow oral directions	1.1.2 Understand social & academic conversations	1.1.3 Understand key ideas of information presented orally	NA	Total
<b>Items</b>	3	6	13		22
<b>Points</b>	3	6	13		22
<b>New Items</b>		6	2		8

New items:

- 2 MC Understand Social Conversations
- 4 MC Understand Academic Conversations

Edit/omit items:

- Omit 2 inference items and rewrite 2 others (#16, 17, 19, 22) - there is no standard for making inferences in Listening

**2008 Form D - Speaking (proposed)**

	2.1.1 Ask & answer questions	2.1.2 Communicate information orally	2.1.3 Plan oral presentations	2.1.4 Deliver oral presentations	NA	Total
<b>Items</b>	6	10	0	2		18
<b>Points</b>	6	12	0	8		26
<b>New Items</b>	3	2				5

New items:

- 2 CR1 Ask Questions
- 2 CR1 Academic-related vocab - similar to item #9 Form C
- 1 CR1 Answer Questions (to replace #4)

IELA Item Development

**2008 Form D1 - Reading (proposed)**

	3.1.1 Text feature s	3.1.2 Graphi c feature s	3.1.3 Phonologic al awareness	3.1.4 Knowledg e of syllables	3.1.5 Decode & determin e meaning of words	3.1.6 Synonym s, antonyms  Multiple meanings	3.1.7 Read with fluency	3.2.1 Reading comp.: Direction s	3.2.2 Reading comp.: Main idea (expositor y)	3.2.3 Reading comp.: draw conclusio ns	3.2.4 Readin g comp.: Fiction (plot, etc.)	NA	Total
<b>Items</b>	1	2	2	0	2	1	1	2	3	1	2	0	17
<b>Points</b>	1	2	2	0	2	1	6	2	3	1	2	0	22
<b>New Items</b>	1					1	1						3

New items:

- 1 MC Text Features (e.g. where would you most likely find a list of chapter headings in a book? a) Table of Contents)
- 1 MC Antonym (e.g., Read the sentence. Which word means the opposite of the underlined word in the sentence?)
- 1 CR6 Fluency

Deleted items:

- 1 MC Fact vs. Opinion

**2008 Form D2 - Reading (proposed)**

	3.1.1 Text feature s	3.1.2 Graphi c feature s	3.1.3 Phonologic al awareness	3.1.4 Knowledg e of syllables	3.1.5 Decode & determin e meaning of words	3.1.6 Synonym s, antonyms  multiple meanings	3.1.7 Read with fluency	3.2.1 Reading comp.: Direction s	3.2.2 Reading comp.: Main idea (expositor y)	3.2.3 Reading comp.: draw conclusio ns	3.2.4 Readin g comp.: Fiction (plot, etc.)	NA	Total
<b>Items</b>	1	1	0	1	2	3	1	2	5	4	3	0	23
<b>Points</b>	1	1	0	1	2	3	6	2	8	5	3	0	32
<b>New Items</b>		1		1		1	1						4

New items:

- 1 MC Graphic Features (e.g., table or graph)
- 1 MC Antonym (e.g., Read the sentence. Which word is an antonym of the underlined word in the sentence?)
- 1 MC Syllables (e.g., How many syllables does the word *pictures* have?)
- 1 CR6 Fluency

Deleted items:

- 1 MC Fact vs. Opinion

IELA Item Development

**2008 Form D1 - Writing (proposed)**

	4.1.1 Writing Process	4.2.1 Write Narrative s	4.2.2 Write Reports	4.3.1 Spell words correctly	4.3.2 Write a variety of sentence types	4.3.3 Capitalizati on & punctuation	4.3.4 Use grammatic al forms	NA	Total
<b>Items</b>	0	1	1	6	2	1	4	0	15
<b>Points</b>	0	2	4	6	4	1	4	0	21
<b>New Items</b>				1	2		4		7

New items:

- 1 MC Spell words correctly
- CR2 Write an interrogative sentence with support (Are you \_\_\_\_\_)
- CR2 Write a simple declarative sentence in response to a picture prompt.
- 1 MC Grammar: identify noun/verb in a sentence
- 2 MC Grammar: choose correct verb form
- 1 MC Grammar: choose correct form of an adjective

**2008 Form D2 - Writing (proposed)**

	4.1.1 Writing Process	4.2.1 Write Narrative s	4.2.2 Write Reports	4.3.1 Spell words correctly	4.3.2 Write a variety of sentence types	4.3.3 Capitalizati on & punctuation	4.3.4 Use grammatic al forms	NA	Total
<b>Items</b>	5*	1	1	4	2	4	5		18
<b>Points</b>	5*	4	4	4	3	4	5		25
<b>New Items</b>					2		4		6

\*4 of these items are also linked to a second standard, therefore they appear twice in the table. They were not, however, counted twice in the Totals.

New items:

- 1 MC Sentence Variety: identify whether a sentence is declarative, interrogative, etc.
- 1 CR2 Sentence Variety: Write a complex sentence, given a sentence frame
- 1 MC Grammar: identify the adjective in a sentence
- 2 MC Grammar: choose correct verb form
- 1 MC Grammar: choose correct form of a superlative adjective (cloze item)

IELA Item Development

**2008 Form E - Listening (proposed)**

	1.1.1 Follow oral directions	1.1.2 Understand social & academic conversations	1.1.3 Understand key ideas of information presented orally	NA	Total
<b>Items</b>	3	6	13	0	22
<b>Points</b>	3	6	13	0	22
<b>New Items</b>		6			6

New items:

- 2 MC Understand Social Conversations
- 4 MC Understand Academic Conversations

Edit/omit items:

- If possible, edit final passage to reduce length. [Reviewers felt strongly about this.]

**2008 Form E - Speaking (proposed)**

	2.1.1 Ask & answer questions	2.1.2 Communicate information orally	2.1.3 Plan oral presentations	2.1.4 Deliver oral presentations	NA	Total
<b>Items</b>	6	10	0	2		18
<b>Points</b>	6	12	0	8		26
<b>New Items</b>	4	2				6

New items:

- 3 CR1 Ask Questions (2.1.1)
- 1 additional CR1 Answer Questions (2.1.1)
- 1 CR1 Academic-related vocab - similar to item #1 (2.1.2)
- 1 CR1 Express feelings (2.1.2)

IELA Item Development

**2008 Form E1 - Reading (proposed)**

	3.1 Reading Process	3.1.1 Text features	3.1.2 Graphic features	3.1.3 Phonological awareness	3.1.4 Decode words using word parts	3.1.5 Use context to determine meaning of words	3.2.1 Reading comp.: Directions	3.2.2 Reading comp.: Main idea (expository)	3.2.3 Reading comp.: draw conclusions	3.2.4 Reading comp.: Fiction (plot, etc.)	NA	Total
<b>Items</b>	1	1	2	1	1	3	2*	3	1	3		17
<b>Points</b>	6	1	2	1	1	3	2*	3	1	3		22
<b>New Items</b>	1	1			1	1			1			5

\*1 item is already counted as a graphic features item. Therefore, do not count twice for Totals.

New items:

- 1 MC Text Features
- 1 MC Decode words with using word parts
- 1 MC inference/draw conclusions item
- 1 MC Vocab in Context item (advanced level)
- 1 CR6 Reading Process (aka Reading Fluency)

Deleted items:

- 1 MC Fact vs. Opinion
- 1 MC Synonym
- 1 MC Read word

**2008 Form E2 - Reading (proposed)**

	3.1 Reading Process	3.1.1 Text features	3.1.2 Graphic features	3.1.3 Phonological awareness	3.1.4 Decode words using word parts	3.1.5 Use context to determine meaning of words	3.2.1 Reading comp.: Directions	3.2.2 Reading comp.: Main idea (expository)	3.2.3 Reading comp.: draw conclusions	3.2.4 Reading comp.: Fiction (plot, etc.)	NA	Total
<b>Items</b>	1	2	1	0	1	4	2	5	3	3		22
<b>Points</b>	6	2	1	0	1	4	2	5	7	3		31
<b>New Items</b>	1	1	1		1	(-3)	1			2		7

## IELA Item Development

New items:

- 1 MC Text Features
- 1 MC Graphic features (graph)
- 1 MC Decode words with using word parts
- 1 MC Directions (to replace poor item)
- 1 CR6 Reading Process (aka Reading Fluency)
- 2 MC Character/Plot items

Deleted items:

- 1 MC Fact vs. Opinion
- 3 MC Vocab in Context (because we have too many)
- 1 MC Directions (poor item)

### 2008 Form E1 - Writing (proposed)

	4.1.1 Writing Process	4.2.1 Write Narrative s	4.2.2 Write Reports	4.3.1 Spell words correctly	4.3.2 Write a variety of sentence types	4.3.3 Capitalizati on & punctuation	4.3.4 Use grammatic al forms	NA	Total
<b>Items</b>	0	1	1	5	2	2	4		15
<b>Points</b>	0	2	4	5	4	2	4		21
<b>New Items</b>		1		1	1		1		4

New items:

- 1 additional MC Spell words correctly
- 1 CR2 Write variety of sentences. (e.g. Write an interrogative sentence with support (Can you...))
- 1 CR2 Write summary of a narrative (similar to D1 #10)
- 1 MC Grammar: identify noun/verb in a sentence

### 2008 Form E2 - Writing (proposed)

	4.1.1 Writing Process	4.2.1 Write Narrative s	4.2.2 Write Reports	4.3.1 Spell words correctly	4.3.2 Write a variety of sentence types	4.3.3 Capitalizati on & punctuation	4.3.4 Use grammatic al forms	NA	Total
<b>Items</b>	5*	1	1	4	3	3	5		17
<b>Points</b>	5*	4	4	4	5	3	5		25
<b>New Items</b>		1			2	1	1		5

\*These are not counted in the Totals because the same items are also counted in 4.3.1 & 4.3.4.

## IELA Item Development

New items:

- 1 CR2 Sentence Variety: Write a complex sentence given a sentence frame.
- 1 MC Sentence Variety: Identify a declarative or interrogative sentence.
- 1 CR4 Write an autobiographical narrative
- 1 MC Grammar: identify a noun/verb/adjective in a sentence
- 1 additional cap & punctuation item

## Appendix C

### Item Writing Overview & Guidelines

In general terms, there are two broad classes of item types: multiple-choice (MC) and open-response (also referred to as constructed-response or CR) items. Issues to consider when writing either MC or CR items for the test are as follows:

- Modality - Is the student listening to the prompt or reading it?
- Proficiency Level - Is it appropriate to ask this question to a student at the targeted proficiency level?
- Grade/Age of Student - Is it appropriate to ask a student at this grade level to answer this question?
- Standard/Objective - Does this item measure what it is supposed to measure?

#### The Prior Knowledge Issue

One thing that you will need to be very aware of when you are writing or editing ELP test items is what knowledge the students should have before taking the test. The purpose of this test is to assess students' *English language proficiency*, not their specific knowledge of American history or geography. When writing multiple-choice items based on a passage, make sure the question can be answered based entirely on the information presented in the passage, without assuming prior knowledge about the topic. On the other hand, also make sure that a student *with* prior knowledge would not be able to answer the question without reading the passage at all.

This same concept holds true for vocabulary words. If a question is designed to determine if a student can derive the meaning of a word from context, then the item should not be asking about a high-frequency word that the student is likely to know.

#### Bias

Potential bias in items or passages is a key issue. There are many different ways that bias can creep into an assessment. The main way to avoid it is for item writers to be sensitive to the potential for bias. Beware of stereotypes and other broad assumptions about gender, race, ethnicity, religion, region, or socioeconomic status.

Regional bias can arise from something as simple as the way that you refer to a soft drink. What is “pop” in the Midwest is “soda” or “cola” in other parts of the country. Using terminology that reflects that state/region for which you are writing is important. An example of socioeconomic bias would be a passage about students at a sleep-away camp. Not all children will have had the opportunity to experience a sleep-away camp first hand and would thus not relate to the passage in the same way other students might. It is likely that the students who have not had that opportunity are from families with lower socioeconomic status. Gender bias may take the form of casting males and females into particular roles or may reflect an imbalance in the content of the entire collection of passages that are used on a test. For instance, if there are 3

## IELA Item Development

passages and 2 are about male scientists and one is about 2 boys playing sports, then girls may not be as interested and hence, their scores may not be an accurate reflection of their ability. Cultural bias is a very important consideration in developing items for English learners, many of whom may have immigrated to this country from a very different culture.

Bias can also take the form of something that potentially will upset students so that they cannot complete the test to the best of their ability. For instance, a passage about a child's pet dying may be upsetting to a child who has recently experienced that grief. It is generally best to avoid topics like death, serious illness, religion, drugs and alcohol, war, and other controversial or potentially upsetting topics. All items will be reviewed for sensitivity and bias, but it is best to avoid these problems when the items are written.

### **Wording**

The wording of items (both MC and CR) is a critical part of item writing. When writing for a test of language ability, it is even more critical. Items must be age-appropriate and also appropriate to the English proficiency level target.

Passages, written to support items, should be written to word count and readability specifications as well as tailored to modality, to grade cluster, and to test level within a grade cluster (i.e., Level 1, Beginner, vs. Level 2, Intermediate/Advanced).

If, in the course of writing items, you need to develop some text (e.g., a conversation, a brief content-area presentation) you should be mindful that the difficulty and the length of the text is grade appropriate as well as proficiency level appropriate. In addition to writing items that are appropriately worded, it is important that the directions given to students be appropriate to their level of English proficiency. If there are directions associated with items that you are writing, those directions should be considered as carefully as the items.

### **Wording Checklist**

- ✓ Is the wording generally appropriate for students at this level?
- ✓ Is the item clear and concise?
- ✓ Has extraneous information been removed from both the stem and the answer choices?
- ✓ Are the answer choices parallel in grammatical structure?
- ✓ Are the answer choices similar in length?
- ✓ If answer choices repeat a word or phrase at the beginning of each choice, has the item been reworked to eliminate that?
- ✓ Does the item avoid using humor and idiomatic expressions?
- ✓ Does the wording reflect the population for whom the test is being developed?
- ✓ Have the words "NOT" and "EXCEPT" been used sparingly or not at all?
- ✓ Has the item been worded positively and does it avoid negative phrasing?
- ✓ Has the central idea been presented in the stem and not in the distractors?

### Multiple-choice Items

All multiple-choice items have the same score weight: 1 point.

The stem is the part of the item that precedes the multiple choice response alternatives. The stem of a multiple choice item should be robust enough so that students know what type of answer they will be looking for before they even look at the answer choices. Look at the following item:

- 00** Pamela likes
- A stories about unicorns.
  - B articles about trains.
  - C\* books about fairies.
  - D tales about mountain climbing.

“Pamela likes” is not a good stem because the student has no idea where the question is going. Conversely, by framing the question with pertinent information, such as in the following example, we can still get at what Pamela likes, but now we have a context. A student conceivably could know that the answer is that Pamela likes books about fairies before looking at the answer choices.

- 00** Pamela likes to read books about
- A unicorns.
  - B trains.
  - C\* fairies.
  - D mountain climbing.

The first stem is just too broad.

Another key issue to keep in mind is that a MC item stem should ask only one concept. Look at the following stem:

- 00** Where did the rabbit go, and how did Wendy know where to look for it?

The main reason that this stem is problematic for multiple choice questions is that it is asking the students to answer two questions. This makes it impossible to determine if a student knew the answer to one part but not the other because a student would get this item wrong by not knowing one of the answers. This is especially an issue in an item such as this one where the two questions require different levels of thought. The first question, “Where did the rabbit go?” is a literal question so students can find the answer directly in the text. The second question, “How did Wendy know where to look for it?” requires inferential thinking and would once again penalize the student who is not as likely to answer inferential questions correctly.

## IELA Item Development

Additionally, since this item requires each answer choice to have two parts, all the options become quite wordy. By splitting the item into 2 separate items as follows, it can be determined if students know either one or both of the pieces of information asked of them.

**00** Where did the rabbit go?

And

**00** How did Wendy know where to look for the rabbit?

Now, the Beginner-level students, who are more likely to miss the second question, will have the opportunity to receive credit for their ability to answer the literal question correctly. Also, by splitting the questions, both items are now simpler and more direct, which benefits all the students.

It is often necessary to contextualize the stem to eliminate more than one correct answer. This is often the case with vocabulary questions. Look at the example below:

**00** The word “spring” means

- A to move suddenly or rapidly upwards or forwards.
- B the season after winter and before summer.
- C an elastic device, typically a spiral metal coil, that can be pressed or pulled but returns to its former shape when released.
- D a place where water wells up from an underground source.

So, what is the correct answer? Without a context, the right answer cannot be determined. Thus, the stem would need to be more specific, as is shown in the following example.

**00** The way that it is used in the passage, the word “spring” means

- A to move suddenly or rapidly upwards or forwards.
- B the season after winter and before summer.
- C an elastic device, typically a spiral metal coil, that can be pressed or pulled but returns to its former shape when released.
- D a place where water wells up from an underground source.

If the passage is about seasons of the year, then B would be the correct answer. If the passage is about components of a mattress then C would be the correct answer. It is often useful to ask vocabulary items in a situation such as this, where the student will need to determine the correct definition based upon the passage or context and not just based upon prior knowledge.

### Distractors

Whereas a strong stem is central to a good item, the distractors, or incorrect answer choices are also important in multiple choice items. Often, the distractors can change the level of difficulty of an item. Look at the following items:

[Text from paragraph 4: Thanks to the oak trees, the ground squirrels had enough nuts to last through the winter season. They did not starve.]

**00** In paragraph 4, what does the word starve mean?

- A glass table
- B flowers
- C birds
- D\* go hungry

**00** In paragraph 4, what does the word starve mean?

- A search for food
- B hibernate
- C run quickly
- D\* go hungry

The first item is much easier (probably too easy) because the reader can immediately eliminate the non-verb choices which would not grammatically fit into the sentence where the target word appears. The reader does not need to look at or understand the wider text context. The second item is more challenging because the response choices are all verb phrases and are all things that ground squirrels might do. The reader must go beyond the single sentence to find the contextual support for the correct answer.

Another key rule when constructing distractors is that there can only be one correct answer. The distractors should be reasonable but not defensible. Look at the distractors in the following example.

**00** At the end of the story, Leon and Linda felt

- A angry
- B\* excited
- C exhausted
- D happy

If excited is the correct answer, then happy is an attractive distractor and should be eliminated. Excited may be the BEST answer but someone who is excited is also happy to a certain extent. Thus, happy could be a defensible choice.

### Parity

It is important that the correct answer not be given away by virtue of its appearance or wording. For instance, in the following example, only one option makes sense grammatically.

- 00** After Sally finished flying her kite, she
- A will make a paper airplane.
  - B takes her sister to the library.
  - C would watch a movie.
  - D\* discovered a bird nest.

D is the only option that grammatically fits. Therefore, an astute student may select this option merely for that reason. The student is, in essence, outsmarting the test and exhibiting good test taking skills, but we do not know if the student knew the answer to the question based on his comprehension of the passage.

Similarly, an answer choice that stands out from the others will draw students to it. This could work either for or against the student, but either way, it is not appropriate test construction. Take a look at the following example, using the same stem:

- 00** After Sally finished flying her kite, she
- A made a paper airplane.
  - B took her sister to the library.
  - C watched a movie and ate popcorn and candy.
  - D\* discovered a bird nest

Option C becomes appealing to some students because it is more detailed than the other options. This option may attract the lower scoring students because there is some appeal to a long answer having the correct information in there somewhere. A better distractor for option C would be, “watched a funny movie.” This would then be about the same length as the other options and a good distractor. Also, if you will notice, all the options now begin with the same verb tense, so the answer is not given away.

### Cueing

When you develop a series of items, it is easy to end up with items that cue one another. This means that the answer to one item can be gleaned from another item. Look at the example below:

- 00** Nick kicked his white soccer ball to
- A Susan.
  - B Jimmy.
  - C Richard.
  - D\* Eliza.

## IELA Item Development

**00** What color is Nick’s soccer ball?

- A pink
- B\* white
- C yellow
- D red

To be able to answer the second question, an astute student would just need to read the previous question to know that Nick’s soccer ball is white. While being able to answer the question by using another question does demonstrate good skills, there is no way to know if the student actually read the passage. In this example, the fix is easy. By removing the word “white” in the first question, the cueing issue is resolved. Although it is often not as easy a fix as it was in this case, it is imperative that item sets be checked for cueing and items be revised if necessary.

### Multiple-Choice Item Checklist

- ✓ Is the item worth asking?
- ✓ Does the MC item ask only one idea?
- ✓ Is the item free of bias?
- ✓ Is the item straightforward? (i.e., not tricky for the target level students)
- ✓ Is the context realistic?
- ✓ Is the difficulty of the item appropriate for the target level?
- ✓ Has all extraneous information in the stem been removed?
- ✓ Does the stem provide enough information?
- ✓ If art is used, does it enhance the item?
- ✓ If art is included to support a Listening passage, is the item still dependent on an understanding of the oral presentation? If the Listening item can be answered based on the art alone, the item needs to be revised.
- ✓ If items are passage dependent, does the student need to read or listen to the selection to successfully answer the question?
- ✓ If the item requires prior knowledge, is that prior knowledge within the parameters of what new immigrant students should be expected to know?
- ✓ If the item asks students to interpret vocabulary, is the vocabulary supported by the context or is it a word or phrase that the target level ELL students should know?
- ✓ Is there only one correct answer?
- ✓ Are all of the distractors plausible?
- ✓ Have tricky distractors been replaced?

## IELA Item Development

- ✓ Are the distractors related to the question?
- ✓ Have common student errors been used as distractors?
- ✓ Are proper names culturally diverse (when there is an opportunity to create the names)?
- ✓ Has a frame of reference been provided if necessary? (i.e., if you are asking about someone's opinion in an article, does the stem note that it is asking about that specific person's opinion)
- ✓ Are any of the answer choices subsets of other answer choices?
- ✓ Does the item match the standard?
- ✓ Is the item appropriate for this grade span and level of fluency?
- ✓ Is the question (stem) simple, direct, and unambiguous?
- ✓ Does the question (stem) use vocabulary and sentence structure that is appropriate to this grade span?
- ✓ Are the response choices reasonably parallel in length and structure?
- ✓ Is there only one clearly correct answer?
- ✓ Is the item free of gender, ethnic, socioeconomic, and regional bias?

### Item Set Checklist

- ✓ Are all of the items unique? (i.e. there are not a lot of very similar items that have virtually the same stem and just different art and/or answers)
- ✓ Have items that cue another item been revised so as to avoid cueing?

### Age-Appropriate Terms

You want to be sure that the language is age appropriate. Using terms that students are not expected to know or that have not been defined will not provide accurate assessment of the students' knowledge.

### Idiomatic Language

Beware of using idiomatic expressions in passages and in item stems unless the specific objective is to test student's knowledge of this particular idiom or their ability to derive the meaning of an idiom from context. Idiomatic expressions make a passage much harder for English language learners (without raising the apparent readability of a passage) and would normally be appropriate only at the Advanced proficiency level. For example, the following

sentence would be inappropriate in a narrative passage aimed at Beginner and Intermediate ELLs: *Shayna was unable to catch the drift of what Rashid was saying.*

Also beware when using multiple-meaning words, especially in passages and items targeted at Beginner-level ELLs. The students may only have learned the most common meaning of the word.

Another difficult concept for ELL students is character name alliteration or rhyme. Students may become lost when reading a passage or an item about Trina, Tina, and Tiny, or about Billy and Willy. Thus, if you have the option of choosing names, select names that will not be confusing, but still reflect our diverse society. It is worthwhile to seek out common names from the minority populations in your state to include in your items.

### **Writing Constructed-Response Items & Scoring Guides**

Whereas multiple-choice items are used to assess the receptive elements of language proficiency (i.e., reading and listening), CR items are most often used to assess the productive elements of language proficiency (i.e., Speaking and Writing).

CR items are characterized in terms of a number of points (1-point, 2-point and 4-point items) and, in addition, as either Short Answer (SA) or Constructed Response (CR). SA items tend to be those with lower point values and with a response that is to a great degree prescribed by the prompt (e.g., What is the item in this picture? *or* What would you say if you wanted to know whether you could check out more than one book?). CR items, on the other hand, tend to be less prescriptive (e.g., tell about an experience, write a persuasive essay).

CR items consist of a spoken and/or written prompt and a scoring guide. Item writers are responsible for creating the prompt and the scoring guides for all 1- and 2-point CR items. The 4-point items use the generic scoring guide that already exists in the IELA.

The scoring guides should include both a descriptor of the point-value response and at least one (preferably more) sample responses.

### **Constructed-Response Items Checklist**

- ✓ Does the item match the standard?
- ✓ Is the item appropriate for this grade span and this fluency level?
- ✓ If the item is based on a passage, does the student, in fact, need to use details from the passage to say or write a response? (If most/many students could answer from prior knowledge, this is not a good item to test reading or listening comprehension.)
- ✓ Is the scope and length of the expected response clear?

## IELA Item Development

- ✓ Has an appropriate scoring guide been created for the item?
- ✓ Is there enough information in the scoring guide to determine how to score different types of responses, including partial responses and grammatically faulty responses?
- ✓ Are sample responses provided for each of the score points greater than 0?