

ENGLISH
LANGUAGE
DEVELOPMENT
ASSESSMENT
K-2

TECHNICAL
MANUAL

Spring 2006



ENGLISH LANGUAGE DEVELOPMENT ASSESSMENT (ELDA)

Developed by

The State Collaborative on Assessment and Student Standards (SCASS)
for Assessing Limited English Proficient Students (LEP)
and
Measurement Incorporated

The ELDA is a product of the collaboration among LEP SCASS member states, the Council of Chief State School Officers (CCSSO), and the U.S. Department of Education. Funding to support development of this assessment has come from LEP SCASS member states, CCSSO, and the U.S. Department of Education through a Section 6112 Enhanced Assessment Instruments grant.

Copyright© 2005 by the Council of Chief State School Officers. All rights reserved.

The Council would like to acknowledge and express its gratitude to the LEP SCASS members [current and former] who have dedicated their time and insights to ensure that ELDA not only meets the requirements of the law, but also meets the needs of students and teachers. LEP SCASS members contributing significantly to the development of ELDA include staff from the following states:

Georgia
Indiana
Iowa
Kentucky

Louisiana
Nebraska
Nevada
New Jersey
Ohio

Oklahoma
South Carolina
Virginia
West Virginia

COUNCIL OF CHIEF STATE SCHOOL OFFICERS

Valerie Woodruff (Delaware), President
Elizabeth Burmaster (Wisconsin), President-Elect
David Driscoll (Massachusetts), Vice-President

G. Thomas Houlihan, Executive Director

Julia Lara, Deputy Executive Director
Division of State Services and Technical Assistance

Council of Chief State School Officers
One Massachusetts Avenue, NW
Suite 700
Washington, DC 20001-1431

Table of Contents

Development and Review	5
Planning	6
Item Development	6
Item Review	7
Field Test Form Development	7
Tryouts	8
Field Test Data Analysis	8
Results	10
Demographics	10
Difficulty	11
Internal Consistency and Validity	17
Cut Scores	21
Evaluations	22
Discussion	26
Difficulty and Discrimination	26
Reliability and Validity	30
Follow-Up	32
References	33
Attachment A: Background Papers	34
Attachment B: Standard Setting Report	72
Attachment C: Evaluation Comments	86

List of Tables and Figures

Tables

1	Demographic Summary	10
2	ELDA K-2 Reading Inventory Summary Statistics: Kindergarten	11
3	ELDA K-2 Reading Inventory Summary Statistics: Grades 1-2	12
4	ELDA K-2 Listening Inventory Summary Statistics: Kindergarten	13
5	ELDA K-2 Listening Inventory Summary Statistics: Grades 1-2	13
6	ELDA K-2 Writing Inventory Summary Statistics: Kindergarten	14
7	ELDA K-2 Writing Inventory Summary Statistics: Grades 1-2	15
8	ELDA K-2 Speaking Inventory Summary Statistics: Kindergarten	16
9	ELDA K-2 Speaking Inventory Summary Statistics: Grades 1-2	16
10	Summary of Row-Total Correlations for Reading Inventory	17
11	Summary of Row-Total Correlations for Listening Inventory	18
12	Summary of Row-Total Correlations for Speaking Inventory	18
13	Summary of Row-Total Correlations for Writing Inventory	19
14	Reliability as a Function of Inventory Length	20
15	Summary of Preliminary Standard Setting: 2005 Field Test Results	21
16	ELDA K-2 Cut Scores	22
17	Summary of Field Test Evaluations	25
18	Comparison of Scores for Native English Speakers and Native Spanish Speakers	29
19	Projected Generalizability Coefficients for Spring 2006 ELDA K-2 Inventories	30

Figures

1	ELDA K-2 Field Test Evaluation Form	23
2	Listening Scores by Grade	26
3	Reading Scores by Grade	27
4	Speaking Scores by Grade	27
5	Writing Scores by Grade	28

The English Language Development Assessment (ELDA) is a product of the joint efforts of the Council of Chief State School Officers (CCSSO) a consortium of the 17 states listed on the inside front cover of this Manual, and Measurement Incorporated (MI). Work on tests for grades 3-12 began shortly after the passage of the No Child Left Behind Act of 2001 (NCLB; P.L. 107-110) and is documented elsewhere. Work on tests for grades kindergarten (K) through 2 began in 2004 and culminated in the development of operational inventories for those grades in the spring of 2006.

The purpose of this Technical Manual is to document the background and technical qualities of those inventories. It is divided into four main sections: Development and Review, Tryout, Results, and Discussion. Like most assessments of its type, there will be ongoing development, and we will continue to document changes and improvements as they occur.

Development and Review

The framework for ELDA K-2 is based on that of the tests for grades 3-12. Frameworks and Standards and Benchmarks were developed by the previous contractor (American Institutes for Research) consortium members, and consultants. Development of the K-2 assessments was guided by these Standards and Benchmarks as well as the general guidelines offered by SCASS LEP and Early Childhood Committee members. Specifically, we designed instruments with the following features in mind:

- Address reporting requirements of NCLB Titles I and III
- Focus on academic language rather than academic content
- Embed items from each content from 3-5 cluster tests
- Reflect developmental states of K-2 students
- Resemble instructional activities as much as possible
- Take no more than about 15 minutes per form

The timeline for initial development and field testing in the spring of 2005 is shown below.

ELDA K-2 Development Schedule		
Event	Begin	End
Organize K-2 development/field test	11/8/2004	11/12/2004
Draft K-2 blueprint	11/11/2004	11/18/2004
Approve K-2 blueprint	11/18/2004	11/22/2004
Draft K-2 test item specifications	11/29/2004	12/10/2004
Approve K-2 item specifications	12/13/2004	12/17/2004
Create K-2 items/activities	12/20/2004	2/18/2005
Review K-2 items/activities	2/22/2005	2/25/2005
Revise K-2 items/activities	2/28/2005	3/14/2005
Approve field test items	3/14/2005	3/18/2005
Assemble field test booklets	3/21/2005	3/25/2005

Approve field test booklets	3/25/2005	3/31/2005
Approve administration manuals	3/25/2005	3/31/2005
Print field test materials	3/31/2005	4/14/2005
Ship field test materials	4/14/05	4/28/05
Conduct field test	5/2/2005	5/31/2005
Collect field test materials	5/31/2005	6/7/2005
Enter field test data; score speaking tapes and writing prompt responses	6/8/2005	7/13/2005
Merge booklet and tape data	7/14/2005	7/15/2005
Analyze field test data	7/18/2005	7/22/2005
Prepare for standard setting	7/25/2005	8/12/2005
Conduct standard setting	8/15/2005	8/20/2005

Planning

Measurement Incorporated representatives Dr. Michael Bunch and Ms. Donna Merritt met with CCSSO staff and consultants to develop an overall plan for the K-2 assessments. These meetings took place in Austin, Texas (December 2004) and Orlando, Florida (January 5, 2005). In addition, we worked with consultants identified by CCSSO to guide the development of the assessments. MI commissioned Dr. Linda Espinoza to draft a concept paper for review by the ELDA K-2 committee at its January meeting in Orlando. From that concept paper and comments from the committee, we constructed inventories.

Item Development

MI invited nine classroom teachers to participate in an item development session in Durham, NC the week of February 14-17, 2005. Working from the concept paper drafted by Dr. Espinoza, the draft recommendation position paper from NAEYC (see Attachment A) and from SCASS Committee members, the teachers worked with Ms. Donna Merritt (MI's director for this project) and Ms. Nancy Rowch (Nebraska) to create inventory entries for Listening, Reading, Speaking, and Writing. The work was collaborative and had as its goal the generation of enough teacher observations/ items to construct three field test inventories.

We adopted the inventory approach over the traditional multiple-choice (MC) and constructed-response (CR) item approach because of the age and developmental stage of the student population (kindergarten through second grade). Each "item" in the initial inventories was a statement regarding an observable student behavior such as the following:

- Follows a two-step instruction in a non-academic setting (e.g., going to the lunchroom)
- Identifies a picture of an object with the same ending sound as 'cat'
- Uses correct English words for manipulatives content, age, and grade appropriate

The item writers used the lists of standards and benchmarks collected from the LEP SCASS member states in the consortium as their guides, along with other materials they brought to the session and those supplied by MI.

The final step in item development was the selection of anchor items from the current tests for grades 3-5. These items were selected on the basis of their relevance to the K-2 content objectives and difficulty indices derived in the summer of 2004. Those items deemed suitable for administration to first and second graders were identified and added to the inventories developed by the item-development committee.

Item Review

CCSSO arranged for Ms. Maria Malagon and Ms. Marjorie Rosenberg to review items developed by the teachers. Ms. Malagon and Ms. Rosenberg reviewed all items for bias and content as they were completed by the item development committee in Durham and reviewed by Ms. Merritt and MI editorial staff. They submitted their recommendations for changes in a formal report to MI. Ms. Merritt incorporated their recommendations into the items and inventories, which she then submitted to CCSSO for review and approval.

We met for a face-to-face review session in CCSSO's offices in Washington, DC on March 11, 2005. This session was similar to those conducted in 2004 for items developed for grades 3-12. At the end of the review session, Ms. Merritt documented all recommendations, made the necessary modifications, and submitted all items to CCSSO for final approval.

Field Test Form Development

Upon approval of the individual entries, Ms. Merritt and Mr. Heavner (Senior Editor at MI) assembled a single field test form for administration in the spring of 2005. The form covered all four content areas and included a sample of the selected anchor items from the current assessments for grades 3-5. We prepared two versions: one for kindergarten only (all entries except the anchor items from grades 3-5) and another for grades 1-2 (all entries, including the grade 3-5 anchor items).

Ms. Merritt submitted completed forms to CCSSO for review and comment. Upon receipt of written comments, she and Mr. Heavner revised the forms as necessary and resubmitted them for final review and approval. Upon approval of the formatted field test forms, Ms. Merritt turned them over to the production team for inclusion in the 2005 field test for grades K-12.

Tryouts

MI staff printed, shipped, and collected field test materials. We barcoded all secure materials for secure delivery and collection. As MI received field test materials from the field, our warehouse staff checked each box for completeness against the outgoing packing list and note any discrepancies. Ms. Jennifer Schaefer followed up with a phone call to resolve any discrepancies.

The student booklets were non-scannable, and there were no scannable answer documents. Therefore, all data in the booklets were hand entered, using double entry to guarantee 100 percent accuracy of entered data. Data entry operators entered the scores for each inventory, using software that prompted the operator to enter student name, demographic information, and school information, and then prompted them from row to row and automatically summed the row scores to derive a total score. This software also locked out invalid (out-of-range) entries, further enhancing the accuracy of the data.

MI scoring leaders conducted training for scorers as they did in 2004, and those scorers evaluated student responses to the writing prompts as well as responses to the speaking prompts, entering scores by barcode number and forwarding them on monitor sheets to information technology (IT) staff. As scorers completed their task, IT staff merged data from the student booklet with score data and the Student Background Questionnaire (SBQ) to produce a final data set for analysis.

Field Test Data Analysis

Dr Michael Bunch and Dr. Kevin Joldersma analyzed the field test data. Given the nature of the data, they conducted the following analyses:

For Inventories

- Score frequency distribution
- Mean and standard deviation by row
- Total score frequency distribution
- Total score mean and standard deviation
- Row-total score correlation
- Tabulation and summary of written comments

For the Merged Data Set

- Correlations between inventory items and SBQ items
- Correlations between inventory items and anchor items
- Cross-tabulation of inventory totals with teacher-assigned levels (from SBQ)
- Preliminary standard setting using the contrasting-groups method

For Evaluation Forms

- Response frequency for each item

- Mean and standard deviation for each item
- Tabulation and summary of written responses

Where there were sufficient numbers of students, we repeated all of these analyses by subgroup (grade, race, sex, and other demographic variables of interest).

Row difficulties and correlations with total score. Dr. Bunch used Excel to calculate item score means, and correlations between item scores and total scores. A second set of correlations (r^2) was produced that removed the effect of each item from the total score to obtain a corrected correlation between item and total score.

Results

Demographics

Table 1 summarizes the demographic composition of the field test sample.

Table 1
Demographic Summary

States	N	%	Ethnicity	N	%
Indiana	402	16.5%	White	150	6.2%
Kentucky	153	6.3%	Black	59	2.4%
Nebraska	251	10.3%	Asian	228	9.4%
New Jersey	756	31.1%	Am.Ind.	2	0.1%
Oklahoma	552	22.7%	Pac.Isl.	10	0.4%
West Virginia	317	13.0%	Hisp/Lat.	1829	75.2%
Total	2431	100.0%	Other	70	2.9%
			MultiEth.	7	0.3%
			Missing	76	3.1%
Grade	N	%	Total	2431	100.0%
K	745	30.6%			
1	831	34.2%			
2	798	32.8%			
Missing	57	2.3%			
Total	2431	100.0%			
First Language	N	%	LEP Status	N	%
Spanish	1784	73.4%	Limited	1743	71.7%
Vietnamese	32	1.3%	Monitored	448	18.4%
Arabic	40	1.6%	Native	106	4.4%
Korean	24	1.0%	Missing	122	5.0%
Chinese	54	2.2%	Other	12	0.5%
Bosnian	10	0.4%	Total	2431	100.0%
Portuguese	1	0.0%			
Somali	5	0.2%			
Urdu	19	0.8%			
Russian	30	1.2%			
Japanese	17	0.7%			
Other	209	8.6%			
Multi	4	0.2%			
Missing	202	8.3%			
Total	2431	100.0%			

Difficulty

Tables 2-9 summarize the difficulty of the individual entries (rows) in the inventories, by grade.

Table 2
ELDA K-2 Reading Inventory Summary Statistics: Kindergarten
Completed Inventories Only

Row	Score				Total	Mean	SD
	0	1	2	3			
1	35	96	107	109	347	1.84	0.98
2	23	54	83	187	347	2.25	0.99
3	67	195	70	15	347	1.10	1.02
4	86	174	66	21	347	1.06	1.01
5	113	173	51	10	347	0.88	1.01
6	138	177	29	3	347	0.70	0.99
7	196	115	35	1	347	0.54	0.99
8	219	110	16	2	347	0.43	0.94
9	77	132	112	26	347	1.25	0.79
10	95	112	123	17	347	1.18	0.82
11	42	123	149	33	347	1.50	0.85
12	4	51	90	202	347	2.41	1.00
13	53	219	58	17	347	1.11	0.99
14	52	175	104	16	347	1.24	0.96
15	10	79	121	137	347	2.11	0.94
16	132	173	37	5	347	0.76	0.96
17	108	166	65	8	347	0.92	0.96
18	258	64	24	1	347	0.33	1.01
19	246	84	16	1	347	0.34	0.91
20	86	117	93	51	347	1.31	0.96
21	86	150	97	14	347	1.11	0.96
22	87	159	95	6	347	1.06	0.84
23	53	185	91	18	347	1.21	0.86
24	80	194	59	14	347	1.02	0.85
25	71	177	91	8	347	1.10	0.84
26	132	185	28	2	347	0.71	0.81
27	221	107	17	2	347	0.42	0.77
28	78	178	79	12	347	1.07	0.77
29	161	136	45	5	347	0.69	0.77
Total						31.67	12.20

Table 3
ELDA K-2 Reading Inventory Summary Statistics: Grades 1 and 2
Completed Inventories Only

		Grade 1								Grade 2							
		Score									Score						
Row	0	1	2	3	Total	Mean	SD	Row	0	1	2	3	Total	Mean	SD		
1	18	74	183	271	546	2.29	0.82	1	5	65	186	295	551	2.40	0.73		
2	17	57	115	357	546	2.49	0.80	2	7	38	100	406	551	2.64	0.67		
3	43	148	231	124	546	1.80	0.88	3	18	131	204	198	551	2.06	0.85		
4	33	85	212	216	546	2.12	0.88	4	14	57	172	308	551	2.40	0.78		
5	45	117	232	152	546	1.90	0.90	5	17	92	194	248	551	2.22	0.83		
6	62	199	217	68	546	1.53	0.85	6	19	139	248	145	551	1.94	0.81		
7	78	134	228	106	546	1.66	0.95	7	22	100	210	219	551	2.14	0.85		
8	69	134	234	109	546	1.70	0.93	8	25	85	232	209	551	2.13	0.84		
9	34	88	224	200	546	2.08	0.88	9	12	90	222	227	551	2.21	0.79		
10	34	143	263	106	546	1.81	0.82	10	9	122	242	178	551	2.07	0.78		
11	26	69	206	245	546	2.23	0.85	11	8	48	160	335	551	2.49	0.72		
12	6	21	62	457	546	2.78	0.56	12	2	12	46	491	551	2.86	0.43		
13	31	193	246	76	546	1.67	0.78	13	12	97	281	161	551	2.07	0.74		
14	31	188	227	100	546	1.73	0.82	14	16	88	255	192	551	2.13	0.78		
15	13	74	177	282	546	2.33	0.80	15	2	38	164	347	551	2.55	0.64		
16	64	198	212	72	546	1.53	0.87	16	18	141	242	150	551	1.95	0.81		
17	54	150	236	106	546	1.72	0.89	17	16	146	248	141	551	1.93	0.80		
18	108	186	212	40	546	1.34	0.88	18	40	152	254	105	551	1.77	0.84		
19	128	206	184	28	546	1.21	0.86	19	39	147	286	79	551	1.74	0.79		
20	44	116	209	177	546	1.95	0.93	20	7	77	192	275	551	2.33	0.76		
21	44	154	256	92	546	1.73	0.84	21	22	116	251	162	551	2.00	0.82		
22	52	139	270	85	546	1.71	0.84	22	19	101	264	167	551	2.05	0.79		
23	37	143	240	126	546	1.83	0.86	23	13	112	221	205	551	2.12	0.81		
24	60	161	210	115	546	1.70	0.92	24	22	108	249	172	551	2.04	0.82		
25	58	147	259	82	546	1.67	0.86	25	27	119	281	124	551	1.91	0.79		
26	79	205	199	63	546	1.45	0.88	26	44	142	266	99	551	1.76	0.84		
27	146	208	142	50	546	1.18	0.93	27	70	156	226	99	551	1.64	0.92		
28	65	161	210	110	546	1.67	0.93	28	27	131	241	152	551	1.94	0.84		
29	121	190	188	47	546	1.29	0.91	29	44	151	239	117	551	1.78	0.87		
Total						52.09	20.00	Total						61.29	18.45		

Table 4
ELDA K-2 Listening Inventory Summary Statistics: Kindergarten
Completed Inventories Only

	Score						
Row	0	1	2	3	Total	Mean	SD
1	15	117	186	265	583	2.20	0.85
2	15	169	225	174	583	1.96	0.83
3	74	231	215	63	583	1.46	0.85
4	96	212	221	54	583	1.40	0.87
5	15	144	183	241	583	2.11	0.87
6	22	146	223	192	583	2.00	0.86
7	83	214	231	55	583	1.44	0.85
Total						12.58	4.99

Table 5
ELDA K-2 Listening Inventory Summary Statistics: Grades 1 and 2
Completed Inventories Only

	Grade 1								Grade 2						
	Score								Score						
Row	0	1	2	3	Total	Mean	SD	Row	0	1	2	3	Total	Mean	SD
1	11	106	223	297	637	2.27	0.80	1	6	50	181	354	591	2.49	0.69
2	9	101	266	261	637	2.22	0.76	2	8	52	222	309	591	2.41	0.71
3	50	158	314	115	637	1.78	0.83	3	21	102	291	177	591	2.06	0.78
4	56	165	322	94	637	1.71	0.82	4	17	116	293	165	591	2.03	0.77
5	10	89	229	309	637	2.31	0.77	5	4	54	156	377	591	2.53	0.69
6	19	112	257	249	637	2.16	0.81	6	10	62	241	278	591	2.33	0.73
7	44	164	323	106	637	1.77	0.81	7	15	106	297	173	591	2.06	0.76
Total						14.22	4.75	Total						15.91	4.42

Table 6
ELDA K-2 Writing Inventory Summary Statistics: Kindergarten
Completed Inventories Only

	Score						
Row	0	1	2	3	Total	Mean	SD
1	10	57	47	39	153	1.75	0.91
2	38	57	39	19	153	1.25	0.97
3	1	45	55	52	153	2.03	0.81
4	23	68	37	25	153	1.42	0.94
5	31	67	44	11	153	1.23	0.85
6	40	75	28	10	153	1.05	0.84
7	40	68	40	5	153	1.07	0.81
8	49	46	35	23	153	1.21	1.06
9	52	60	39	2	153	0.94	0.80
10	3	13	31	106	153	2.57	0.73
11	2	51	75	25	153	1.80	0.72
12	19	42	51	41	153	1.75	0.99
13	6	82	41	24	153	1.54	0.80
14	39	44	49	21	153	1.34	1.01
15	50	55	40	8	153	1.04	0.90
16	37	34	64	18	153	1.41	0.98
Total						23.41	10.65

Table 7
ELDA K-2 Writing Inventory Summary Statistics: Grades 1 and 2
Completed Inventories Only

Grade 1								Grade 2							
Score								Score							
Row	0	1	2	3	Total	Mean	SD	Row	0	1	2	3	Total	Mean	SD
1	11	101	254	139	505	2.03	0.75	1	11	66	238	240	555	2.27	0.75
2	35	91	262	117	505	1.91	0.83	2	13	69	244	229	555	2.24	0.76
3	4	49	124	328	505	2.54	0.70	3	4	27	75	449	555	2.75	0.58
4	21	85	212	187	505	2.12	0.83	4	10	58	172	315	555	2.43	0.75
5	37	113	275	80	505	1.79	0.79	5	10	81	282	182	555	2.15	0.72
6	38	140	239	88	505	1.75	0.83	6	16	101	264	174	555	2.07	0.78
7	47	91	307	60	505	1.75	0.78	7	14	54	386	101	555	2.03	0.62
8	37	57	240	171	505	2.08	0.86	8	12	54	222	267	555	2.34	0.74
9	58	50	349	48	505	1.77	0.77	9	19	32	410	94	555	2.04	0.60
10	3	13	65	424	505	2.80	0.50	10	2	10	48	495	555	2.87	0.42
11	4	62	247	192	505	2.24	0.69	11	3	28	207	317	555	2.51	0.62
12	12	62	129	302	505	2.43	0.80	12	5	30	82	438	555	2.72	0.60
13	5	81	227	192	505	2.20	0.74	13	10	39	169	337	555	2.50	0.71
14	20	100	223	162	505	2.04	0.82	14	11	78	224	242	555	2.26	0.77
15	54	33	297	121	505	1.96	0.86	15	33	46	285	191	555	2.14	0.80
16	42	30	288	145	505	2.06	0.82	16	12	30	316	197	555	2.26	0.66
Total						33.47	9.55	Total						37.58	8.37

Table 8
ELDA K-2 Speaking Inventory Summary Statistics: Kindergarten
Completed Inventories Only

Row	Score				Total	Mean	SD
	0	1	2	3			
1	6	26	32	50	114	2.11	0.93
2	7	17	40	50	114	2.17	0.90
3	5	19	43	47	114	2.16	0.86
4	6	13	42	53	114	2.25	0.86
5	6	21	50	37	114	2.04	0.85
6	4	20	46	44	114	2.14	0.83
7	12	12	37	53	114	2.15	0.99
8	7	16	61	30	114	2.00	0.81
9	7	20	69	18	114	1.86	0.75
10	4	17	67	26	114	2.01	0.72
11	7	20	61	26	114	1.93	0.81
12	10	21	51	32	114	1.92	0.90
Total						24.72	8.98

Table 9
ELDA K-2 Speaking Inventory Summary Statistics: Grades 1 and 2
Completed Inventories Only

Row	Grade 1							Row	Grade 2						
	Score				Total	Mean	SD		Score				Total	Mean	SD
0	1	2	3	0				1	2	3					
1	3	27	63	91	184	2.32	0.78	1	3	8	71	141	223	2.57	0.63
2	6	13	84	81	184	2.30	0.74	2	3	11	81	128	223	2.50	0.66
3	5	21	72	86	184	2.30	0.78	3	3	14	77	129	223	2.49	0.68
4	3	12	69	100	184	2.45	0.69	4	3	5	73	142	223	2.59	0.61
5	4	19	91	70	184	2.23	0.72	5	3	5	100	115	223	2.47	0.61
6	3	18	85	78	184	2.29	0.71	6	3	4	89	127	223	2.52	0.61
7	1	10	76	97	184	2.46	0.63	7	1	8	70	144	223	2.60	0.58
8	4	20	102	58	184	2.16	0.70	8	3	12	97	111	223	2.42	0.66
9	8	15	120	41	184	2.05	0.69	9	3	12	118	90	223	2.32	0.64
10	5	15	120	44	184	2.10	0.65	10	3	6	119	95	223	2.37	0.61
11	6	22	119	37	184	2.02	0.67	11	3	10	111	99	223	2.37	0.64
12	15	29	95	45	184	1.92	0.85	12	17	18	94	94	223	2.19	0.88
Total						26.61	7.16	Total						29.41	6.49

Internal Consistency and Validity

Tables 10-13 summarize the correlations between individual inventory row scores and total scores. The first two entries refer to the correlation with total score, r1 indicating the correlation without correction (i.e., with that item included in the total) and r2 indicating the corrected correlation (i.e., with that item removed from the total). The final index (rRate) is the correlation between score and teacher rating.

Table 10
Summary of Row-Total Correlations for Reading Inventory: All Students

Row	r1	r2	rRate
1	0.73	0.71	0.56
2	0.67	0.64	0.49
3	0.84	0.83	0.61
4	0.86	0.85	0.59
5	0.88	0.87	0.63
6	0.87	0.85	0.60
7	0.86	0.85	0.57
8	0.87	0.85	0.58
9	0.85	0.84	0.63
10	0.78	0.76	0.54
11	0.81	0.79	0.54
12	0.58	0.56	0.39
13	0.77	0.75	0.46
14	0.81	0.80	0.54
15	0.69	0.67	0.52
16	0.85	0.83	0.56
17	0.87	0.86	0.61
18	0.86	0.85	0.53
19	0.85	0.84	0.49
20	0.80	0.79	0.56
21	0.85	0.84	0.57
22	0.87	0.86	0.58
23	0.84	0.83	0.59
24	0.86	0.84	0.57
25	0.84	0.83	0.60
26	0.85	0.84	0.54
27	0.80	0.78	0.51
28	0.80	0.79	0.58
29	0.78	0.76	0.48
Total			0.68

Table 11
Summary of Row-Total Correlations for Listening Inventory: All Students

Row	r1	r2	rRate
1	0.79	0.71	0.42
2	0.86	0.81	0.48
3	0.87	0.82	0.50
4	0.88	0.82	0.50
5	0.86	0.80	0.50
6	0.86	0.81	0.51
7	0.87	0.82	0.51
Total			0.57

Table 12
Summary of Row-Total Correlations for Speaking Inventory: All Students

Row	r1	r2	rRate
1	0.84	0.81	0.59
2	0.87	0.85	0.65
3	0.88	0.86	0.64
4	0.85	0.82	0.58
5	0.89	0.87	0.62
6	0.87	0.85	0.58
7	0.72	0.67	0.46
8	0.90	0.87	0.61
9	0.86	0.83	0.57
10	0.87	0.85	0.56
11	0.88	0.85	0.58
12	0.78	0.73	0.52
Total			0.68

Table 13
Summary of Row-Total Correlations for Writing Inventory: All Students

Row	r1	r2	rRate
1	0.76	0.72	0.48
2	0.81	0.77	0.47
3	0.83	0.81	0.48
4	0.86	0.83	0.53
5	0.83	0.81	0.53
6	0.84	0.81	0.56
7	0.81	0.78	0.50
8	0.86	0.83	0.47
9	0.79	0.75	0.44
10	0.52	0.48	0.24
11	0.76	0.73	0.47
12	0.76	0.72	0.35
13	0.80	0.77	0.41
14	0.84	0.81	0.57
15	0.79	0.76	0.43
16	0.78	0.75	0.43
Total	1.00		0.58

Explanation of Entries for Tables 2-13

- Each observable behavior takes up one row in the inventory.
- Scores range from 0 (no mastery) to 3 (complete mastery)
- Row = Row number in the inventory
- Total (column heading) = Total number of students for whom we received completed inventories (all rows filled in)
- Mean = average score for the row across all students observed
- SD = standard deviation of all entries for this row
- r1 = correlation between scores for this row and score for the total inventory
- r2 = correlation between scores for this row and score for the total inventory minus this row (corrected correlation)
- rRate = correlation between scores for this row and teacher rating (1 to 5)
- Total (row heading at bottom of table) refers to total test statistics

We performed generalizability analyses (Brennan, 1983) to determine generalizability coefficients (equivalent to coefficient alpha or KR-20) and phi coefficients, lower bound estimates of reliability equivalent to coefficient KR-21. Table 14 gives the reliability coefficients for all four inventories. The two reliability indices are shown as rel1 for the generalizability coefficient and rel2 for the phi coefficient. From the outset, we had envisioned inventories of various lengths. The values for the inventories as field tested (e.g., 29 rows for Reading) are

shown in **bold**. The values for the inventories as implemented in the spring of 2006 are shown in *italics*. In the case of Listening, the field-tested length and the operational length are the same (7 rows). In all cases, the generalizability and phi coefficients for inventories, as field tested and as operationally tested, were .90 or above.

Table 14
Reliability as a Function of Inventory Length

Reading				Listening				Writing				Speaking		
Rows	Rel1	Rel2		Rows	Rel1	Rel2		Rows	Rel1	Rel2		Rows	Rel1	Rel2
2	0.79	0.72		1	0.69	0.62		2	0.75	0.68		1	0.69	0.67
4	0.88	0.84		2	0.81	0.76		4	0.86	0.81		2	0.82	0.80
6	0.92	0.88		3	0.87	0.83		6	0.90	0.86		3	0.87	0.86
8	0.94	0.91		4	0.90	0.87		8	0.92	0.89		4	0.90	0.89
10	0.95	0.93		5	0.92	0.89		9	<i>0.93</i>	<i>0.90</i>		5	0.92	0.91
12	0.96	0.94		6	0.93	0.91		10	0.94	0.91		6	0.93	0.93
14	<i>0.96</i>	<i>0.95</i>		7	0.94	0.92		12	0.95	0.93		7	0.94	0.94
16	0.97	0.95		8	0.95	0.93		14	0.95	0.94		8	<i>0.95</i>	<i>0.94</i>
18	0.97	0.96		9	0.95	0.94		16	0.96	0.94		9	0.95	0.95
20	0.97	0.96		10	0.96	0.94		18	0.96	0.95		10	0.96	0.95
22	0.98	0.97		11	0.96	0.95		20	0.97	0.95		11	0.96	0.96
24	0.98	0.97		12	0.96	0.95		22	0.97	0.96		12	0.96	0.96
26	0.98	0.97		13	0.97	0.95		24	0.97	0.96		13	0.97	0.96
28	0.98	0.97		14	0.97	0.96		26	0.98	0.96		14	0.97	0.97
29	0.98	0.97		15	0.97	0.96		28	0.98	0.97		15	0.97	0.97
30	0.98	0.97						30	0.98	0.97		16	0.97	0.97
32	0.98	0.98						32	0.98	0.97		17	0.97	0.97
34	0.98	0.98										18	0.98	0.97
36	0.99	0.98										19	0.98	0.98
38	0.99	0.98										20	0.98	0.98
40	0.99	0.98										21	0.98	0.98
												22	0.98	0.98
												24	0.98	0.98

Cut Scores

Initial standard setting: 2005. Each teacher administering the K-2 inventories in 2005 was asked to rate the proficiency of each student on a scale of 1 to 5 (5 being Fully English Proficient) on each component. The standard application of this procedure (cf. Jaeger, 1989) is to compile the raw scores for all examinees classified by their teachers in Level 1, Level 2, and so on. Cut scores can then be derived in a number of ways. The simplest is to find the midpoint between two means or medians. For example, if the mean raw score for all students classified by their teachers as Level 1 were 20 and the mean raw score for all students classified by their teachers as Level 2 were 30, the cut score for entry into Level 2 would be 25, the midpoint between 20 and 30. Table 15 summarizes the contrasting-groups standard setting results

Table 15
Summary of Preliminary Standard Setting: 2005 Field Test Results
(Groups based on teacher rating of overall proficiency)

Reading (87 Points)			
Level	N	Mean	Cut
1	211	23.0	
2	367	44.2	33.6
3	353	53.7	48.9
4	298	67.3	60.5
5	136	69.7	68.5
Total	1365		
Listening (21 Points)			
Level	N	Mean	Cut
1	135	6.8	
2	298	10.5	8.6
3	533	14.1	12.3
4	471	16.9	15.5
5	268	18.3	17.6
Total	1705		
Writing (48 Points)			
Level	N	Mean	Cut
1	171	21.5	
2	335	31.9	26.7
3	313	37.2	34.5
4	231	41.3	39.2
5	89	40.0	40.6
Total	1139		

Table 15 Continued

Speaking (36 Points)			
Level	N	Mean	Cut
1	27	10.9	
2	62	21.7	16.3
3	149	26.1	23.9
4	137	30.6	28.3
5	119	32.6	31.6
Total	494		

Level	Label
1	Prefunctional
2	Beginning
3	Intermediate
4	Advanced
5	Fully English Proficient (FEP)
Cut	Preliminary cut score based on contrasting-groups method

Follow-up standard setting. A second standard setting was conducted in January 2006. Panelists were ELL educators and other stakeholders in five participating states: Iowa, Louisiana, Nebraska, Ohio, and South Carolina. A total of 16 individuals participated (9 for the Kindergarten inventories and 7 for the Grades 1-2 inventories). A complete report is included in Attachment B. Results are summarized in Table 16.

**Table 16
ELDA K-2 Cut Scores**

Level	Listening (Out of 21 Points)		Reading (Out of 42 Points)		Speaking (Out of 24 Points)		Writing (Out of 27 Points)	
	K	1-2	K	1-2	K	1-2	K	1-2
2	4	6	8	10	6	8	7	8
3	9	11	20	22	12	13	16	17
4	15	16	36	31	18	18	21	21
5	19	19	40	39	22	22	26	25

Evaluations

Teachers administering the ELDA K-2 inventories in the spring of 2005 completed background questionnaires about their students and evaluation forms regarding the inventories. Figure 1 shows the evaluation form we asked teachers to complete. Table 17 summarizes their evaluations of the inventories. Written comments are summarized in Attachment C.

Figure 1
ELDA K-2 Field Test Evaluation Form

Please take a few minutes to complete this evaluation form. Your comments and suggestions will be helpful to us in making the final version of the ELDA for students in grades K-2 an accurate measure of your students' language development.

Respond to the following statements by checking the appropriate box.

Evaluation Statements		Strongly Agree	Agree	Disagree	Strongly Disagree
1	I received the materials I needed in time to prepare to administer the field test.	SA	A	D	SD
2	The field test materials were well labeled and easy to use.	SA	A	D	SD
3	The test items in the first part of the Student Booklet were appropriate for my students.	SA	A	D	SD
4	The directions for administering the items in the first part of the Student Booklet were easy to understand and follow.	SA	A	D	SD
5	The tasks on the Reading Inventory were appropriate for my students.	SA	A	D	SD
6	The directions for administering the Reading Inventory were easy to understand and follow.	SA	A	D	SD
7	The tasks on the Listening Inventory were appropriate for my students.	SA	A	D	SD
8	The directions for administering the Listening Inventory were easy to understand and follow.	SA	A	D	SD
9	The tasks on the Writing Inventory were appropriate for my students.	SA	A	D	SD
10	The directions for administering the Listening Inventory were easy to understand and follow.	SA	A	D	SD
11	The tasks on the Speaking Inventory were appropriate for my students.	SA	A	D	SD
12	The directions for administering the Speaking Inventory were easy to understand and follow.	SA	A	D	SD
13	Directions for completing the Student Background Questionnaire were easy to understand and follow.	SA	A	D	SD
14	I had enough time to administer the test items in the Student Booklet to all my students.	SA	A	D	SD
15	I had enough time to complete the Inventories for all my students.	SA	A	D	SD
16	I was able to observe my students long enough to make meaningful entries in the Inventories.	SA	A	D	SD

Figure 1 Continued

Which items, if any, in the first part of the Student Booklet were inappropriate for most of your students? If you did not administer one of these tests, leave that row blank.

Inventory	Inappropriate Item(s)	Comments
Reading		
Listening		
Writing		
Speaking		

Please describe your experience with the Inventories, using the table below. If you did not administer a particular Inventory, leave that row blank.

Inventory	Rows that were difficult to score	Rows that were inappropriate for my students	Comments
Reading			
Listening			
Writing			
Speaking			

General Comments

Table 17
Summary of Field Test Evaluations

Scores by Count					
Question #	Total # Scores	Strongly Agree	Agree	Disagree	Strongly Disagree
1	108	11	31	34	32
2	109	14	77	13	5
3	93	4	52	26	11
4	101	16	62	21	2
5	104	6	54	32	12
6	107	11	65	26	5
7	104	6	69	19	10
8	103	8	69	22	4
9	102	4	59	27	12
10	103	7	75	16	5
11	103	7	70	21	5
12	98	7	61	24	6
13	100	9	62	23	6
14	101	8	44	30	19
15	105	5	22	41	37
16	102	6	37	30	29
Scores by Percentage					
Question #	Total # Scores	Strongly Agree	Agree	Disagree	Strongly Disagree
1	108	10	29	31	30
2	109	13	71	12	05
3	93	04	56	28	12
4	101	16	61	21	02
5	104	06	52	31	12
6	107	10	61	24	05
7	104	06	66	18	10
8	103	08	67	21	04
9	102	04	58	26	12
10	103	07	73	16	05
11	103	07	68	20	05
12	98	07	62	24	06
13	100	09	62	23	06
14	101	08	44	30	19
15	105	05	21	39	35
16	102	06	36	29	28
Average	103	08	55	25	12

Discussion

The ELDA K-2 inventories were developed within a framework that is consistent with the framework for the current ELDA instruments for grades 3-12. The inventory approach was deemed by the consortium members, developers, and consultant reviewers as the most appropriate approach for young English language learners. The results of the field tests, conducted in the spring of 2005, show that the inventory approach worked well, that students were able to respond to the entries, and that scores on the inventories accurately discriminated among groups of students identified by teachers. Key features of the inventories are summarized below.

Difficulty and Discrimination

A review of Tables 2-9 shows that the items were generally appropriately difficult for this population. Out of 64 inventory entries, only four (all on the Reading inventory) had raw score means below 0.5 (for kindergarten students), and only eleven had raw scores above 2.5 (three each in Reading and Writing, one in Listening, and four in Speaking, all for grades 1-2). Moreover, the entries discriminated among students in predictable and desirable ways. Second graders scored higher than first graders, and first graders scored higher than kindergartners. Figures 2-5 summarize the differences among grades.

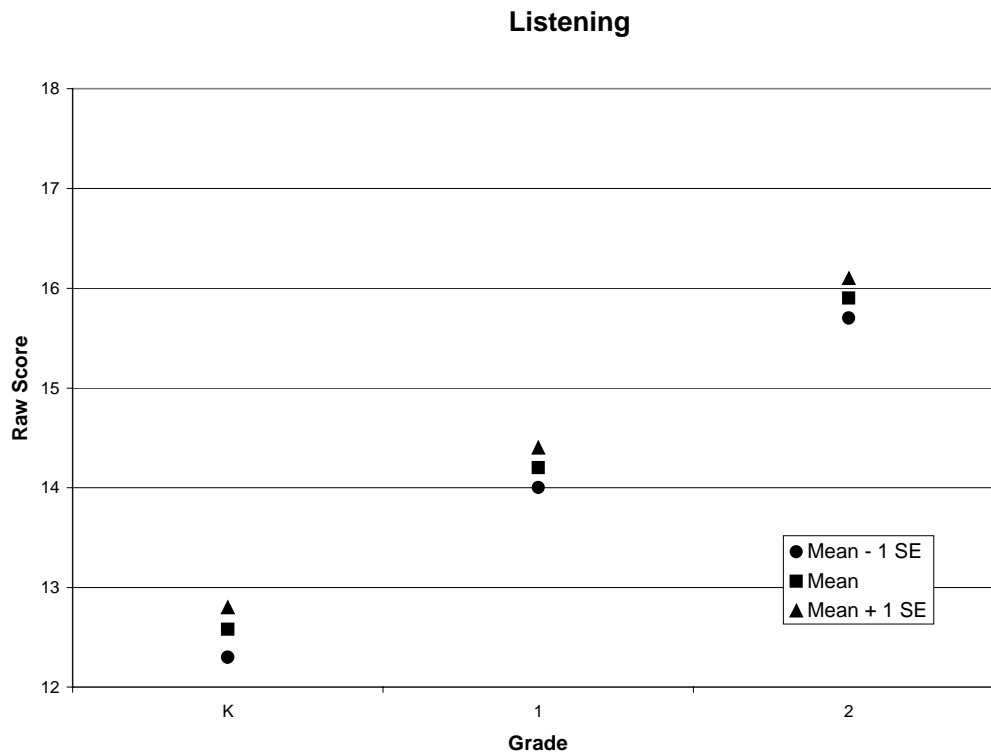


Figure 2. Listening Scores by Grade

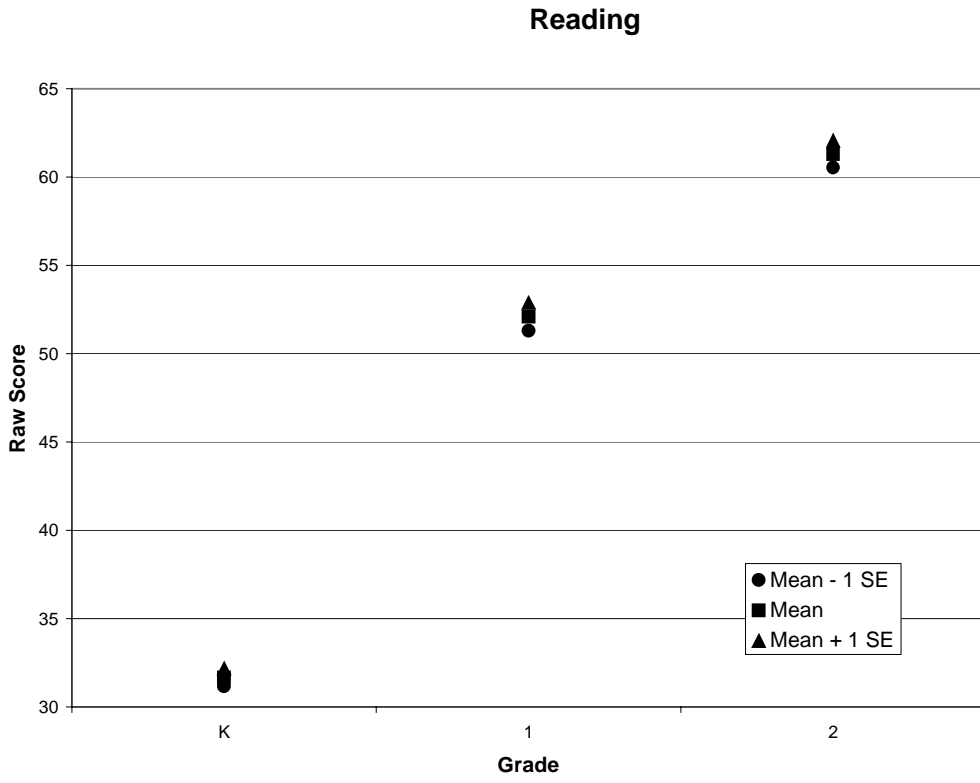


Figure 3. Reading Scores by Grade

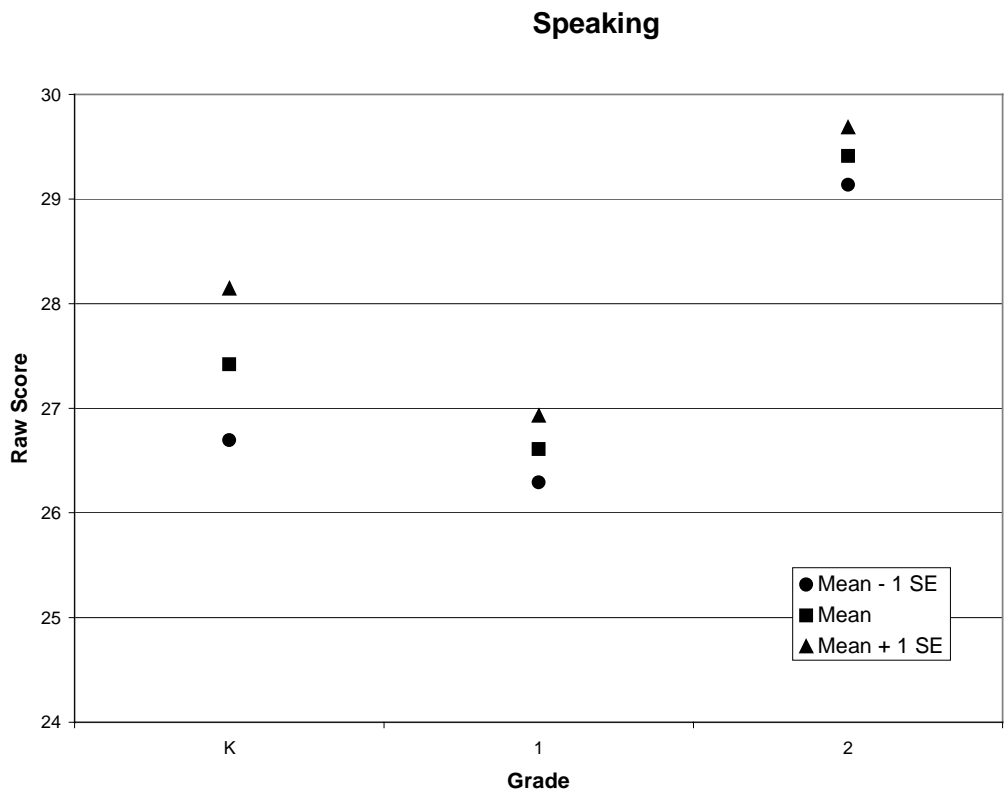


Figure 4. Speaking Scores by Grade

Writing

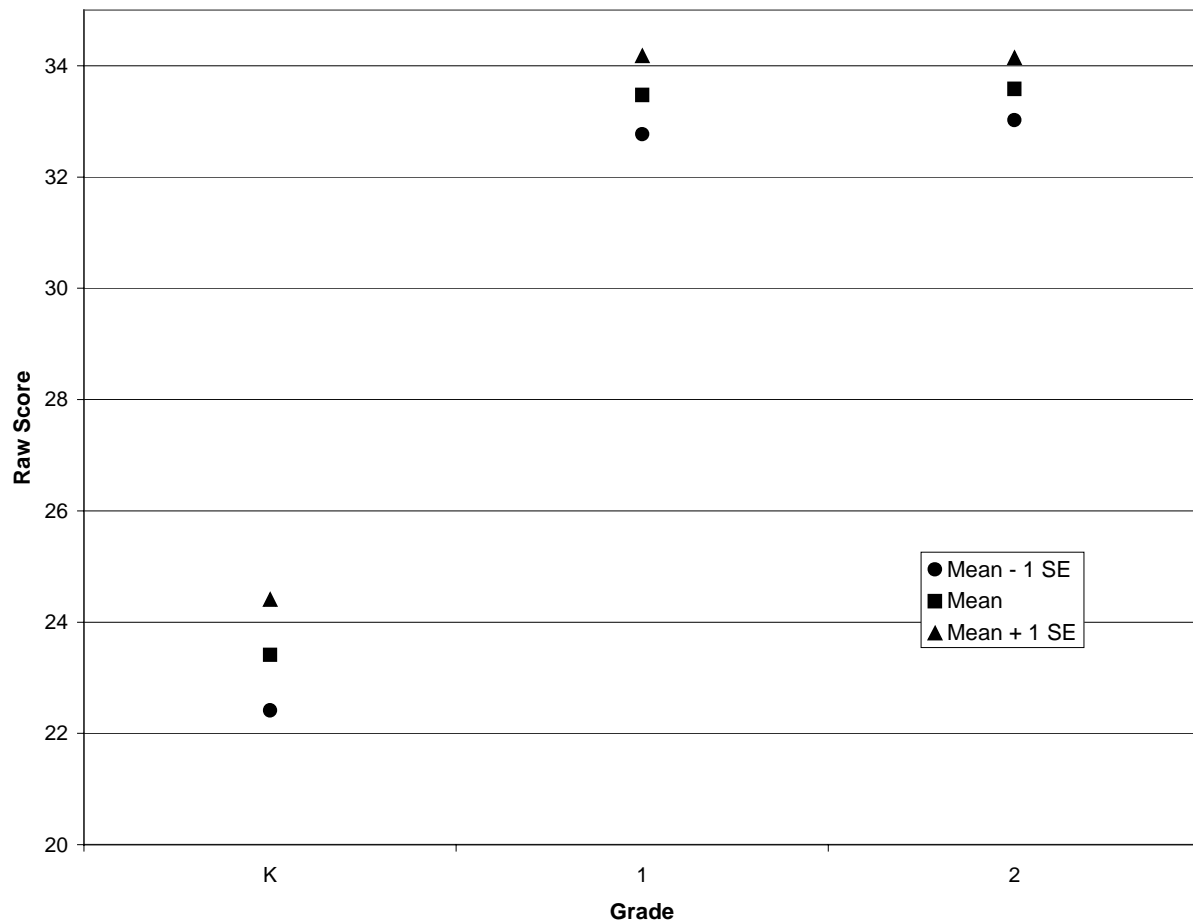


Figure 5. Writing Scores by Grade

In Figures 2-5, the mean score for each grade is represented by a black square. Scores one standard error of the mean above and below the mean are represented by black circles (below) and triangles (above). Note that for Listening and Reading in particular, there is absolutely no overlap of the score distributions from one grade to the next. For example, the Mean + 1 SE score for Reading for grade K is much lower than the Mean – 1 SE score for grade 1, and the Mean + 1 SE for grade 1 is much lower than the Mean – 1 SE for grade 2.

Speaking and Writing present different pictures of achievement by grade. In Speaking, the kindergartners actually have a slightly higher mean than the first graders (27.4 vs. 26.6). Figure 4 shows that these two score distributions overlap, so the difference is not significant. By grade 2, however, students have clearly separated themselves from students in grades K and 1. Here we see no score overlap with lower grades. In Writing, the differences are a bit more subtle. From Table 5, it would appear that kindergartners do not write at all; their scores are so far below those of students in grades 1 and 2. At first glance, it appears that students in grades 1 and 2 do not differ from each other at all. Indeed, the means are very nearly identical (33.5 for grade 1 and 33.6 for grade 2). The distinction is in the lower end of the achievement scale. The score

for Mean – 1 SE for grade 2 is slightly higher for grade 2 than for grade 1, indicating that by the end of second grade, nearly everyone is learning to write.

Because differences among these groups may be attributed to developmental differences rather than differences in language acquisition, we also compared performances of native English speakers to those of the largest comparison group we could find, students whose first language was Spanish. Table 18 summarizes these differences. In every case, native English speakers significantly outscored non-native speakers on each test.

Table 18
Comparison of Scores for Native English Speakers and Native Spanish Speakers

Reading (87 points)				Speaking (36 points)			
Native English Speakers		Spanish First Language		Native English Speakers		Spanish First Language	
Level	N	Level	N	Level	N	Level	N
1	0	1	174	1	0	1	21
2	8	2	292	2	0	2	38
3	11	3	276	3	3	3	95
4	12	4	182	4	8	4	87
5	24	5	84	5	18	5	68
Mean	69.8	Mean	48.8	Mean	32.3	Mean	26.5
SD	15.41	SD	21.47	SD	0.69	SD	7.88
N	55		1008	N	29		309
t		9.59	p<.01	t		12.51	p<.01
Listening (21 points)				Writing (48 points)			
Native English Speakers		Spanish First Language		Native English Speakers		Spanish First Language	
Level	N	Level	N	Level	N	Level	N
1	0	1	103	1	2	1	70
2	1	2	213	2	29	2	70
3	9	3	420	3	27	3	95
4	16	4	336	4	0	4	87
5	39	5	184	5	0	5	68
Mean	17.2	Mean	13.8	Mean	58.0	Mean	26.5
SD	3.26	SD	1.73	SD	0.69	SD	7.88
N	65		1256	N	58		390
t		8.28	p<.01	t		77.02	p<.01

As Table 18 shows, all t values are significant at the .01 level, indicating clear distinctions between native English speakers and other students. While these differences could also reflect differences in familiarity with school environments and test-taking strategies, they are consistent enough, both internally and relative to other measures of achievement, to indicate that the inventories are measuring important language acquisition content standards.

Reliability and Validity

Reliability. Tables 10-13 summarized the item-total correlations for the inventories, while Table 14 summarized the total test reliability for each inventory. As noted in that section, all reliability coefficients, both for the inventories as field tested (long version) and as set for operational administration in the spring of 2006 (short version), are .90 or above. Even the lower-bound estimates of the inventories are .90 or above, indicating that the inventories are extremely internally consistent. Reliability coefficients for the shortened inventories to be administered in Spring 2006 are summarized in Table 19.

Table 19
Projected Generalizability Coefficients for Spring 2006 ELDA K-2 Inventories

Inventory	Rows	Generalizability Coefficient
Reading	14	.96
Listening	7	.94
Speaking	8	.95
Writing	9	.93

Given the fact that they are relatively short and open to some interpretation by the teachers completing them, these reliability coefficients are quite high. As with all statewide exams, however, all statistical properties of tests being administered in any given year are based on field test statistics and will be recomputed after the tests have been given in their operational form. Inasmuch as the operational versions of the inventories follow the same format as the field-test versions (with changes in overall length and some fine-tuning of wording of some entries), we expect operational statistics to deviate very little from the field-test statistics.

Validity. There was a time when the validity of a test was expressed simply as the correlation between scores on that test and scores on some other test or external criterion. At least since Samuel Messick's panoramic rethinking of the concept of validity, very few test developers have taken such a simplistic approach to validity. Messick (1989) defined validity as an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the *adequacy* and *appropriateness* of *inferences* and *actions* based on test scores or other modes of assessment. (p. 13; emphasis in the original). To be valid for this purpose, tests must 1) measure what is taught in schools, 2) provide a standard that is consistent with what students need to know and be able to do, 3) provide for consistent measurement of that standard over time, and 4) be free of bias. In addition, the tests should demonstrate a reasonable degree of association with other measures of school achievement. The validity of the tests, therefore, is to be established over a period of time through the accumulation of a body of evidence that clearly demonstrates that all these criteria have been met. This same point of view is also reflected in the most recent version of *Standards for Educational and Psychological Testing*, jointly published by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education (1999) and followed by most test developers.

To date, we have gathered a significant amount of information about the validity of the ELDA K-2 inventories, relative to points 1-4 above. Evidence is summarized below.

1. Measure what is taught in schools

As pointed out on pages 4-6 of this Manual, the ELDA K-2 inventories are based on a collection of English language acquisition content standards developed by the states participating in the consortium with the cooperation of national consultants and test development contractors and in a manner consistent with the latest research on language acquisition in young children (see Attachment A). At each stage of the development process, classroom teachers and other stakeholders familiar with the student population and content reviewed blueprints, test items (inventory entries), support materials, and manuals. As the inventories were field tested, teachers administering them were given opportunities to provide feedback on the relevance of the inventories to their classroom instruction. As the Evaluation section of this Manual shows, there was a good fit, according to most teachers.

2. Provide a standard that is consistent with what students need to know and be able to do.

The performance level descriptors (PLDs) included in Attachment B represent a collective effort of CCSSO, the participating states, and content expert consultants. These PLDs are very similar to PLDs submitted to classroom teachers who participated in the 2005 field tests. Those teachers endorsed the PLDs by applying them and sorting their students into five performance levels in a way that was generally consistent with the performances of students on the inventories. After the spring 2006 administration of the inventories, the ratings of classroom teachers, once again sorting their students into the five levels defined by the PLDs included in Attachment B, will provide evidence of the validity of the use of the inventories for assigning students to achievement levels. The plan is for CCSSO to take teacher ratings into consideration in modifying cut scores as necessary to better reflect the reality of the classroom.

3. Provide for consistent measurement of that standard over time.

The ELDA K-2 inventories were field tested in the spring of 2005. Spring 2006 will be the first operational administration of the inventories. The content standards are well defined and clearly aligned with test items and performance level descriptors (and even preliminary cut scores). The blueprints and test construction guidelines are well established. Procedures are in place to use the embedded items from the 3-5 ELDA to calibrate both the 2005 and 2006 versions of the inventories so that overall test difficulty remains comparable and cut scores retain meaning over time. As with any new program, consistent measurement over time remains more of a promise than a reality. However, given the attention to detail and involvement of such a wide array of stakeholders and content experts, the promise appears to be one that will be kept.

4. Be free of bias.

At each stage of development, bias review was an integral part of overall item and test review. Any item or inventory entry suspected of bias was revised or rejected. In the field test of 2005, the performances of males and females and various language groups were compared. Reviewers

found no bias by gender or by first language. It should be pointed out, however, that for approximately 75 percent of all tested students Spanish was listed as the first language. Numbers for other languages were too small to permit meaningful comparisons.

Association with other measures.

As noted above, performance on ELDA K-2 inventories correlates well with teacher judgments of student performance. At grades K-2, there is very little other evidence of performance, and what we have from teachers is very encouraging. Correlations between teacher ratings and scores on the Reading and Speaking inventories were both .68, while correlations for Listening and Writing were somewhat lower (.57 for Listening and .58 for Writing). While the correlations for Reading and Speaking are quite good, there is room for improvement in the other two. For Listening, the correlation is fairly consistent across grades (.51 for K, .58 for grade 1, and .57 for grade 2). For Writing, the situation is different; correlations for grades 1 and 2 are nearly as high as those for Reading and Speaking (.65 for both grades). The correlation between Writing inventory performance and teacher ratings for kindergarten students is only .34, reinforcing the previous comment that many kindergartners probably do not write at all. Even this relatively low correlation serves to bolster the case for validity: At grades where more writing would be expected, the correlation between ELDA K-2 inventory scores and teacher ratings of writing achievement are higher.

Follow-Up

Subsequent to the 2005 field test, there were two key meetings concerning ELDA K-2. The first was in Savannah, Georgia, on July 6-7, 2005. At this meeting, state representatives presented many of the concerns voiced by K-2 teachers regarding the length and difficulty of administering the inventories. In August, Dr. Dina Castro joined the MI team of developers to begin revising the inventories with two key goals: shorten the inventories, and provide more support for teachers who administer them. On December 7-9, 2005, member state representatives met again in Washington, DC, to review revised materials. These materials were ultimately approved with modifications during December 2005 and January 2006. The final materials were submitted and approved on January 31, 2006 and will be administered to students in the spring of 2006. As those inventories are administered, a new round of analyses will be completed and this Manual updated.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for Educational and Psychological Testing*. Washington, DC: Author.
- Brennan, R. L. (1983). *Elements of Generalizability*. Iowa City, IA: American College Testing Program.
- Hays, W. L. (1988). *Statistics* (4th Ed.), Fort Worth: Holt, Rinehart, and Winston, Inc.
- Jaeger, R. M. (1989). Certification of student competence. In R. L. Linn (Ed.), *Educational Measurement* (3rd Ed.). New York: American Council on Education & Macmillan Publishing Co.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (3rd Ed.). New York: American Council on Education & Macmillan Publishing Co.
- Wright, B. D. & Stone, M. H. (1979). *Best Test Design*. Chicago: MESA Press.
- Wright, B. D. & Masters, G. N. (1982). *Rating Scale Analysis*. Chicago: Mesa Press.

Attachment A Background Papers

- **Castro & Bunch**
- **NAEYC**

Using ELDA With Students in Grades K-2

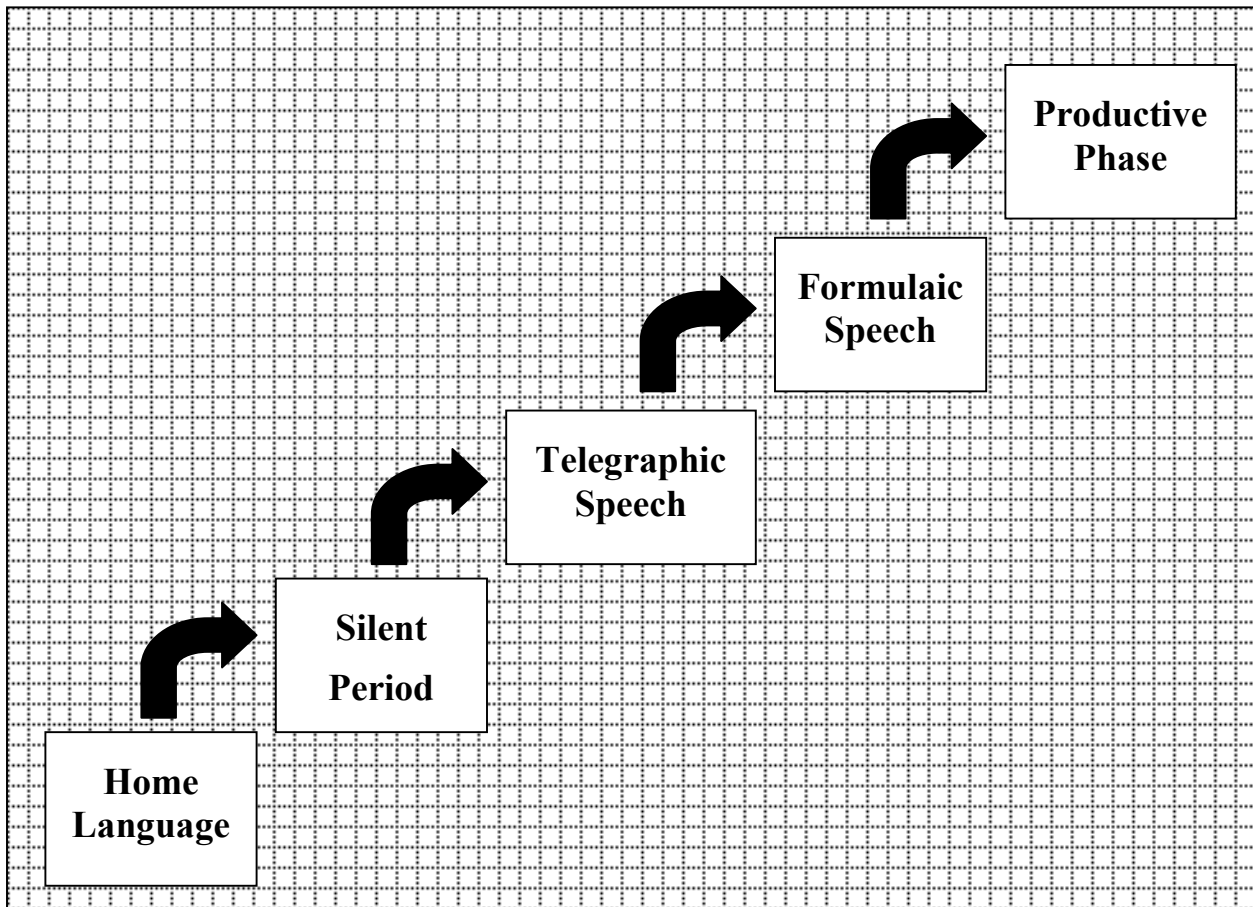
Dina Castro (Frank Porter Graham Child Development Center)

Michael B. Bunch (Measurement Incorporated)

Before using the ELDA with children who are English Language Learners (ELLs), it is important to understand how second language acquisition occurs in grades K-2. This information will help you in both the administration and interpretation of the assessments.

How do young children learn a second language?

Children most often learn a new language in a fixed pattern:



When exposed for the first time to an environment in which you do not understand the language spoken, a natural reaction is to try to communicate verbally using the only language you know. This may have happened to you if you have ever traveled to a non-English speaking country, and did not know the language spoken there. This is precisely what ELLs would generally do when

they first attend school in the United States, **using the home language** in their first attempts to communicate with teachers and classmates. At some point, they will realize that their home language does not help them communicate in that environment and will stop talking. This has been called a non-verbal or **silent period**.

During the non-verbal period, ELLs are learning English by listening and observing everything happening around them. It is very important to engage ELLs in all classroom activities, even if they are not speaking in English yet, because they will be using every opportunity to learn their second language. After some time, ELLs will start speaking English, first using one or two-word phrases that will help them communicate basic needs (e.g., bathroom, lunch) and get into social interactions (e.g., want play, mine). They will be using what is often called **telegraphic speech**.

Little by little children will start using more complex language structures. First they will use memorized phrases (e.g., “I want to play”, “I don’t like it” “I don’t know”). This is the **formulaic speech** phase of second language acquisition. During this phase, it is tempting to assume that an ELL is already fully proficient in English, since the child may be using grammatically correct expressions in English. However, what it may be that the child is using memorized expressions instead of creating original sentences. Eventually, ELLs will start constructing their own sentences getting into a **productive** phase in their second language acquisition process. Children will gradually improve in their use of English grammatical structures (syntax), increase the number of words they know in English (vocabulary), learn how to pronounce them correctly (phonology) and learn the appropriate use of English expressions in different situations (pragmatics).

What are the factors that influence second language acquisition?

Learning does not happen in the same way for everyone, there are individual factors that may influence the learning process, and second language learning is not an exception. The child’s age, personality, motivation and learning style are related to the time it takes someone to learn a second language and the level of proficiency achieved. Regarding age, research shows that older children have more advanced cognitive and social skills that will help them in learning a second language. While younger children will have the advantage of acquiring a native-like pronunciation when speaking English, older children will have the previous knowledge from their first language to help them understand more complex language functions and concepts. This is particularly relevant for literacy related skills.

In terms of personality, children who are more outgoing and are looking for opportunities to interact socially with other children and adults may have more opportunities to learn a second language than children who are shy and don’t feel comfortable interacting with many people. Motivation is another factor that may influence second language learning. Usually, the interest in making friends and the desire to feel part of the group act as strong motivators for children to learn a second language. With regard to learning styles, for some children learning will be facilitated when visual stimuli are used along with words to illustrate and idea or concept, while for other children learning is easier when they have the opportunity to manipulate objects and observe events.

How does literacy development occur in English language learners?

Children who are English language learners are more likely to become readers and writers of English when they are already familiar with the concepts in their home language. Among the language skills that are more closely related to literacy, **phonological awareness** has been found to be one of the strongest predictors of the speed and efficiency of reading acquisition. Phonological awareness is the ability to attend to the sound structure of words. In the case of ELLs, research has shown that phonological awareness transfers from the first language to the second language.

Alphabet knowledge, concepts of print, and syntactic knowledge may also transfer from the first language to the second language, but only if these skills have been developed sufficiently in the primary language. This is particularly relevant for children whose home language uses the alphabetic system, since the letters will be the same although they may not be pronounced with the same sounds. The conventions of reading will be the same (e.g., reading from left to right, knowing that text, rather than pictures, carries the message) and the awareness of how text is used.

How do I interpret ELLs' performance on the ELDA inventories?

The ELDA inventories are designed to measure the level of English proficiency in the areas of speaking, listening, reading and writing. Depending on age and previous experience, ELLs may have more advanced levels of proficiency in speaking, listening, reading and writing in their home language than in English. However, those abilities are not included in the ELDA. Thus, **the ELDA is a tool to determine English proficiency level and not developmental levels on any of these areas.** Not performing at grade level on the ELDA will be an indicator of lower level of proficiency in English, but not necessarily an indicator of a developmental delay or disability. **Results of the ELDA should not be used to determine eligibility for special education services.**

Finally, be aware that becoming fully proficient in a second language takes several years. ELLs in grades K-2 will first learn enough English to get into social interactions, maintain a conversation and communicate basic needs, but the proficiency level needed to understand more complex instructions and academic concepts will likely take longer to learn. When given instructions verbally without visual support or prompting, ELLs who are at an early stage of second language acquisition may fail to give the correct answer, not for lack of content knowledge but rather for not being able to understand the instruction. To be sure that you are making the correct judgment about children's performance, do not base your score on a one-time observation. It is recommended that you observe children over several days and in different settings.

Bibliography

- August, D., & Hakuta, K. (Eds.). (1998). *Educating language-minority children*. National Academy of Sciences- National Research Council, Board on Children, Youth and Families, and Institute of Medicine. Washington, D.C.: National Academy Press.
- Cisero, C. A., & Royer, J. M. (1995). The development and cross-language transfer of phonological awareness. *Contemporary Educational Psychology*, 20, 275-303.
- Galambos, S. J. & Goldin-Meadow, S. (1990). The effects of learning two languages on levels of metalinguistic awareness. *Cognition*, 34 (1), 1-56.
- Gregory, E. (1996). *Making Sense of a New World*. London, UK: Paul Chapman Publishing Ltd.
- Kamii, C., Long, R., Manning, M., & Manning, G. (1990). Spelling in kindergarten: A constructivist analysis comparing Spanish-speaking and English-speaking children. *Journal of Research in Childhood Education*, 4(2), 91-97.
- Kenner, C., & Gregory, E. (2003). Becoming biliterate. In N. Hall, J. Larson, & J. Marsh (Eds.), *Handbook of early childhood literacy* (pp. 178-188). Great Britain, UK: Sage Publications.
- Lee, P. (1996) Cognitive development in bilingual children: A case for bilingual instruction in early childhood education. *The Bilingual Research Journal*, 20(3), 499-522.
- Lopez, L. M. & Greenfield, D. B. (2004) The cross-language transfer of phonological skills of Hispanic Head Start children. *Bilingual Research Journal*, 28(1).
- McLaughlin, B. (1992). *Myths and misconceptions about second language learning: What every teacher needs to unlearn*. Educational Practice Report 5. National Center for Research on Cultural Diversity and Second Language Learning. Retrieved June 14, 2005, from <http://www.ncela.gwu.edu/pubs/ncrcdsl/epr5.htm>
- Nagy, W., McClure, E., & Mir, M. (1997). Linguistic transfer and the use of context by Spanish-English bilinguals. *Applied Psycholinguistics*, 18, 431-452.
- Tabors, P. O. (1997). *One child, two languages*. Baltimore, MD: Paul H Brookes.
- Tabors, P. O. (1998) What early childhood teachers need to know: Developing effective programs for linguistically and culturally diverse children and families. *Young children*, 53(6), 20-26.

**Screening and Assessment
of Young English-Language Learners**
**Supplement to the NAEYC Position Statement on Early Childhood
Curriculum, Assessment and Program Evaluation**
Adopted Summer 2005

Available at: http://www.naeyc.org/about/positions/ell_supplement.asp

Introduction

The National Association for the Education of Young Children (NAEYC) and the National Association of Early Childhood Specialists in State Departments of Education (NAECS/SDE) in 2003 published the joint position statement "Early Childhood Curriculum, Assessment, and Program Evaluation: Building an Effective, Accountable System in Programs for Children Birth through Age 8." The position statement explains what effective assessment looks like for all young children.

One of the indicators of good assessment is that it is linguistically and culturally responsive for all children, including children whose home language is not English. The aim of this document, which was requested by experts in the field, is to explain and expand on the meaning of "linguistically and culturally responsive"; to discuss other issues uniquely related to the screening and assessment of young English-language learners; 1 and to make specific recommendations to increase the probability that all young English-language learners will have the benefit of appropriate, effective assessment of their learning and development. All aspects of the full position statement are important and relevant for young English-language learners, and readers of this document should first read the curriculum, assessment, and program evaluation position statement (NAEYC & NAECS/SDE 2003), bearing in mind that this document serves as a supplement to the full position statement.

This supplement is intended for a range of audiences in the early childhood profession who have a stake in the well-being of young English-language learners, including education decision makers and policy makers, program administrators, teacher educators, researchers, and current and future early childhood educators. Because it is intended for a wide audience, this document is not a how-to manual offering detailed technical advice.² The absence of detailed technical advice here is not meant to abnegate the need for or diminish the importance of such support. Rather, it is hoped that readers will use this document for different purposes: to better articulate their own philosophies, needs, and challenges in this area; to create or revise policies and practices; to guide the development of more resources; and to develop a forward-looking vision of how to improve the development and education of young English-language learners.

Why Now?

A number of factors make the need for this document especially urgent, not the least of which is the dramatic rise in ethnic diversity in the United States. Citizens from diverse racial and ethnic groups now comprise about one-third of the U.S. population. Hispanics are the largest minority population; there are approximately 40 million people of Hispanic origin living in the United States, from Mexico, Central and South America, Puerto Rico, Cuba, and other places. In 2003, Hispanic children from birth to age 5 represented 21 percent (4.2 million) of the total number of children in that age range (19.8 million) (Collins & Ribeiro 2004). In the public school setting alone, there are more than two million English-language learners in prekindergarten through grade 3 (Abedi, Hofstetter, & Lord 2004). Although Spanish accounts for almost 80 percent of the non-English languages (Abedi, Hofstetter, & Lord 2004), more than 460 languages are spoken by English-language learners nationwide (Hepburn 2004).

Because early childhood professionals are serving so many more young English-language learners, there is a great need for appropriate and effective assessment to support these children's learning and development. The field lacks the kinds of assessment tools and well-trained professionals required to implement effective assessment practices for this group. This gap has consequences for the children: without appropriate ways to assess young English-language learners, teachers cannot make the best decisions about how and what to teach, because they are unable to capture a full and accurate picture of children's interests, abilities, and learning needs. Also, the lack of good tools and practices can lead to underidentification of children who have special needs, resulting in the failure to provide needed services.

Simultaneously, problems with the assessment of young English-language learners sometimes lead to overidentification of special needs—that is, misdiagnosing language delays and other learning and developmental disabilities—resulting in children being taken out of the classroom to receive services they do not need and thus missing out on other beneficial activities (Gutierrez-Clellen & Kreiter 2003; NAEYC 1995). Compounding these immediate difficulties are the enduring danger, stigma, and frustration that result when children are mislabeled, which is especially grievous with vulnerable children and families who already must cope with multiple challenges. For these and other reasons, it is critical that the early childhood field improve its ability to screen, assess, and effectively use the results of assessments with young English-language learners.

The Right to be Assessed

Whether enrolled in a child care center, Head Start program, family child care setting, or public school, the millions of young English-language learners in the United States have the right to experience ongoing, effective assessment that supports their learning and development. Observing and documenting the progress of young children is central to the practice of early childhood professionals. Through individual assessments, teachers can appreciate children's unique qualities and talents and individualize instruction (Hyson 2003; NAEYC & NAECS/SDE 2003); make decisions about classroom activities, such as what books to read and what instructional strategies to use; identify children who might benefit from special services; and

have more informed communication with families and with other professionals (McAfee, Leong, & Bodrova 2004). Through program evaluation and accountability assessments, decision makers can make improvements in programs and services that benefit children. Therefore, when children are not regularly and appropriately assessed, they miss out on an effective education: they may fail to receive beneficial special services; classroom activities may not be effectively individualized; and important data that can lead to broader program improvement may be lacking.

Young English-language learners have the right to be assessed for the same reasons and benefits as all children. Moreover, they have the right to be assessed with high-quality assessments and under assessment conditions responsive to the needs of young English-language learners. NAEYC's belief in the right of children to be assessed stems from research and professional values.³

Acknowledging the Challenges

Because assessment is key in determining effective practices and enhancing program quality, it is of great concern when real-world obstacles stand in the way of appropriately assessing young English-language learners. The biggest challenge, of course, is the scarcity of appropriate assessments to use with young English-language learners. For the vast majority of the hundreds of languages represented in the United States, there simply are no assessments. For some languages, such as Spanish, assessments exist, but many of them do not meet technical standards for reliability and validity, or they contain culturally unfamiliar material or are predicated on culturally inappropriate expectations for children's responses. Moreover, even when high-quality assessments are available, programs rarely have qualified bilingual staff to assess children in their home language.

Other obstacles include lack of financial resources, lack of articulated program philosophies or mission statements about English-language learners, difficulty attracting and retaining bilingual and bicultural staff, lack of community awareness about the importance of the issue, and lack of professional development opportunities, to name a few. Later sections of this document will propose steps toward addressing these challenges.

These conditions make it difficult to implement recommendations or improve policies and practices for the assessment of young English-language learners. NAEYC recognizes the gap between realities faced in the field and the vision conveyed in these recommendations. Nonetheless, it is hoped that the recommendations will help policy makers, program administrators and supervisors, assessment specialists, advocates, and practitioners to know what to strive for and how to begin to create environments for improved assessment of young English-language learners.

Overview of Recommendations

Recommendations about the screening and assessment of young English-language learners are presented in seven categories:

1. appropriate uses of screening and assessment;
2. culturally and linguistically appropriate assessment;
3. characteristics of assessments used to improve instruction;
4. use of standardized formal assessments;
5. characteristics of those conducting assessments;
6. the role of family; and
7. needs in the field.

Each category, or section, begins with a general recommendation and then lists and discusses several indicators of effective practices and policies.

The seven recommendations and their indicators are outlined on pages 4-5. An expanded section follows, giving the rationales for the recommendations and discussing the indicators.

1. Using Screening and Assessment for Appropriate Purposes. *As with assessment of all young children, assessment of young English-language learners should be guided by specific, beneficial purposes, with appropriate adaptations to meet the needs of children whose home language is not English.*

1a. Screening: Young English-language learners are regularly screened using linguistically and culturally appropriate screening tools. Results of screenings are used to determine what further supports and services are needed.

1b. Assessment to promote learning: Assessments of young English-language learners are used primarily to understand and improve children's learning; to track, monitor, and support development in all areas, including language development; and to identify disabilities or other special needs.

1c. Program evaluation and accountability: Young English-language learners are included in program evaluation and accountability systems, and culturally and linguistically appropriate assessment instruments and procedures are used. Inclusion of English-language learners in accountability systems never acts as a disincentive for programs to serve English-language learners.

2. Culturally and Linguistically Appropriate Assessments. *In assessing young English-language learners, great emphasis should be given to the alignment of assessment tools and procedures with the specific cultural and linguistic characteristics of the children being assessed.*

2a. All screenings and assessments used with young English-language learners are culturally appropriate.

2b. All screenings and assessments used with young English-language learners are linguistically appropriate.

2c. Translations of English-language instruments are carefully reviewed for linguistic and cultural appropriateness by native speakers well versed in the complex issues of assessment and translation.

3. Characteristics of Assessments Used to Improve Instruction. *The primary purpose of assessing young English-language learners should be to help programs support their learning and development; classroom-based assessment should maximize the value of the results for teachers' curriculum planning and teaching strategies.*

3a. Programs rely on systematic observational assessments, using culturally and linguistically appropriate tools as the primary source of guidance to inform instruction and to improve outcomes for young English-language learners.

3b. Assessments for young English-language learners are based on multiple methods and measures.

3c. Assessments for young English-language learners are ongoing; special attention is given to repeated assessments of language development over time.

3d. Assessments for young English-language learners involve two or more people.

3e. Assessments for young English-language learners are age appropriate.

4. Using Standardized Formal Assessments. *The development of state and other accountability systems has led to increased use of standardized formal assessments of young children. Specific considerations about the development and interpretation of these assessments should guide their use with young English-language learners.*

4a. It is appropriate to use standardized formal assessments to identify disabilities or other special needs, and for program evaluation and accountability purposes. They may also contribute to monitoring and improving learning at an individual level as part of a more comprehensive approach to the assessment of young English-language learners.

4b. Decision makers and those conducting assessments are aware of the concerns and cautions associated with using standardized formal assessments with young English-language learners.

4c. Decision makers and test developers carefully attend to test development issues, including equivalence and norming.

4d. Decision makers and those conducting assessments know appropriate conditions for using and interpreting standardized formal assessments with young English-language learners.

5. Characteristics of Those Conducting Assessments. *Whatever the purpose of the assessment, those conducting assessments of young English-language learners should have cultural and linguistic competence, knowledge of the children being assessed, and specific assessment-related knowledge and skills.*

5a. It is primarily teachers who assess young English-language learners, but paraprofessionals, assessment assistants, and specialized consultants also play an important role.

5b. Those assessing young English-language learners are bilingual and bicultural.

5c. Those assessing young English-language learners know the child.

5d. Those assessing young English-language learners are knowledgeable about language acquisition, including second-language acquisition.

5e. Those assessing young English-language learners are trained in and knowledgeable about assessment in general and about considerations in the assessment of young English-language learners in particular.

6. The Role of Family in the Assessment of Young English-Language Learners. *Families of young English-language learners should play critical roles in the assessment process, being closely involved in a variety of appropriate ways.*

6a. Professionals involved in the assessment of young English-language learners seek information and insight from family members in selecting, conducting, and interpreting assessments.

6b. Programs refrain from using family members to conduct formal assessments, interpret during formal assessments, or draw assessment conclusions.

6c. Professionals involved in assessment regularly inform and update families on their child's assessment results in a way that is easily understood and meaningful.

7. Needs in the Field. *Resources should be invested to ensure rapid progress on several fronts: expanding the knowledge base; developing more and better assessments; increasing the number of bilingual and bicultural professionals; and creating professional development opportunities for administrators, supervisors, practitioners, and other stakeholders in effective assessment of young English-language learners.*

7a. Scholars provide an expanded knowledge base about second-language acquisition and the development of young English-language learners.

7b. More and better assessments are developed to meet the most pressing needs.

7c. Policy makers, institutions of higher education, and programs adopt policies and practices to recruit and retain a diverse early childhood workforce, with a focus on increasing the number of bilingual and bicultural early childhood professionals.

7d. Early childhood professionals, including program administrators, receive ongoing opportunities for professional development and support in the area of assessing young English-language learners. ©

RECOMMENDATIONS AND INDICATORS, WITH RATIONALES

1. Using Screening and Assessment for Appropriate Purposes

RECOMMENDATION

As with assessment of all young children, assessment of young English-language learners should be guided by specific, beneficial purposes, with appropriate adaptations to meet the needs of children whose home language is not English.

Assessment of young children should occur for specific and beneficial purposes (NAEYC & NAECS/SDE 2003). The purpose of each assessment must be clear to those conducting the assessments, program administrators, and policy makers or other decision makers who review and use the results, and assessments results should be used only for the purpose for which the assessment was designed (Scott-Little, Kagan, & Clifford 2003). Because so few appropriate assessments for young English-language learners are available, this caution is especially pertinent in the assessment of these children.

INDICATORS

1a. Screening: Young English-language learners are regularly screened using linguistically and culturally appropriate screening tools. Results of screenings are used to determine what further supports and services are needed.

For all children, screening usually entails a brief, standardized procedure that can quickly determine whether a child may have a problem in some area that would require further assessment and possibly special services to address the problem (McAfee, Leong, & Bodrova 2004; NAEYC & NAECS/SDE 2003). As with all children, young English-language learners should receive regular screenings. Screenings for young English-language learners should be used with two ends in mind: (a) to detect a possible problem in areas including health and physical development, social and emotional development, and cognitive development and (b) to detect a possible problem in the area of language development, including first- and second-language acquisition.

What should differentiate screening of young English-language learners from the screening of monolingual English-speaking children are the tools used and the patterns of follow-up after the screenings. Screenings should use linguistically and culturally appropriate tools (see section 3) that meet appropriate technical standards. Screenings should occur in the child's home language and English, if the child speaks some English, and screeners should accept a child's use of code-switching (i.e., using words and grammar rules from both languages).

Follow-up after screening is critical. If a potential problem is detected, further in-depth assessment with specialists should be scheduled to determine whether the problem exists, and if so, how best to address it (McAfee, Leong, & Bodrova 2004). Because young English-language learners show variable paths to language development and because there is limited research regarding what levels of language proficiency should be expected (Gutiérrez-Clellan & Kreiter 2003), it can be difficult to interpret the results of language screening for individual children. When screening results suggest that follow-up is needed, it is important to involve a specialist who can communicate with the child and family in the child's home language and has specialized expertise in the relevant area of diagnostic assessment.

1b. Assessment to promote learning: Assessments of young English-language learners are used primarily to understand and improve children's learning; to track, monitor, and support development in all areas, including language development; and to identify disabilities or other special needs.

As with all young children, assessment of young English-language learners should be used primarily to understand and promote a child's learning and development as well as to respond to concerns raised by screenings. Specifically, assessment of young English-language learners should be used to (a) guide curriculum planning, teaching strategies, and the provision of learning opportunities in all areas (further discussed in section 3); (b) monitor development and learning in all domains-including children's content knowledge, skills, and capabilities; (c) determine language proficiency and ongoing language development in both the child's home language and English, as appropriate; and (d) identify children with developmental disabilities or delays, emotional impairments, physical disabilities, and other conditions that indicate the need for special services.

1c. Program evaluation and accountability: Young English-language learners are included in program evaluation and accountability systems, and culturally and linguistically appropriate assessment instruments and procedures are used. Inclusion of English-language learners in accountability systems never acts as a disincentive for programs to serve English-language learners.

The use of carefully designed evaluations that hold a program accountable for producing positive results can benefit all children, including young English-language learners, and the use of child-level assessment results for program evaluation and accountability purposes has become more prevalent in recent years-in Head Start programs, for example, as well as in public school settings. However, program-level evaluations often are attached to high stakes, such as decisions about funding for the program. Therefore, when child-level assessments are used in accountability systems, they should be subject to particularly rigorous standards for their design, instrumentation, and analysis (NAEYC & NAECS/SDE 2003; Scott-Little, Kagan, & Clifford 2003).

As noted earlier, young English-language learners have the right to be assessed for all of the reasons all young children are assessed, and young English-language learners should be included in program evaluations and tracking systems so their progress as a group may be monitored and services improved.⁴ Every effort should be made to find appropriate instruments so that these

children can be included. At present, however, very few assessments developed for young English-language learners meet the rigorous standards necessary for use as part of program evaluation and accountability. When it is the case that appropriate assessment instruments and procedures are not available for children who are not proficient in English, these children should not be included in program-evaluation or accountability procedures, but test developers, program administrators, and policy makers should rapidly work to find ways to include them by developing or supporting the development of appropriate assessments.

In large-scale accountability systems, assessments typically rely on standardized formal instruments that directly assess young children through questions and answers or, with older children, written responses. (Recommendations and indicators related to these types of assessments are found in section 4.) In addition to developing more appropriate and effective formal standardized instruments, policy makers and educators should proactively seek ways to include English-language learners' results from other types of assessments, such as observation-based assessments.

Two primary audiences rely on results from program evaluations-program administrators (for decision-making purposes about curriculum, staffing, etc.) and policy makers (for accountability purposes)-and both audiences should examine results for young English-language learners in order to track their development over time as a group. The first purpose of program-level evaluation is for program administrators (e.g., directors, principals) to gather information to guide planning and decision making for their program. Administrators should examine progress in children's home language and English, and in other major domains of learning and development (e.g., social-emotional skills, mathematical thinking, the arts). With this kind of information, they can determine the effects of various approaches to teaching English-language learners and answer questions about curriculum, staffing, approaches to using English and other languages for instructional purposes, ideal groupings of children by language background or proficiency, and so forth. Results from this type of evaluation are also important for communicating with families, the community, and policy makers about the efficacy of a program.

A second purpose of program-level evaluation is accountability at the local, state, or federal level; that is, to provide evidence to entities funding the programs and to the community that programs are meeting determined goals and providing expected services. Policy makers and others look closely at results from these assessments, and English-language learners should be included in order to ensure that these children make progress and that programs receive the support they need to serve young English-language learners.

It is important to ensure that the inclusion of young English-language learners in accountability systems does not discourage programs from serving these children. Administrators who fear that results from young English-language learners' assessments will reflect negatively on their program might limit or even deny services to these children. Policy makers should use this assessment information to create incentives for programs to serve and promote progress in the development of young English-language learners.

Program evaluation for accountability purposes requires that information be gathered from large numbers of children. As recommended in NAEYC and NAECS/SDE's joint position statement on curriculum, assessment, and program evaluation (2003), sampling (assessing only a representative percentage of children) is the most efficient and effective means of capturing data for accountability purposes in a way that is both scientifically rigorous and sensitive to program needs. Administrators and policy makers should include enough English-language learners in their sampling plans to permit conclusions to be reached about the probable effectiveness of the strategies being used to support young English-language learners and the programs serving these children and their families.

In addition to improving program performance and services, results from these types of evaluations also will, in the long run, allow early childhood professionals to create a better picture of the trajectories of young English-language learners as they experience different kinds of early education and as they move through the primary grades.

2. Culturally and Linguistically Appropriate Assessments

RECOMMENDATION

In assessing young English-language learners, great emphasis should be given to the alignment of assessment tools and procedures with the specific cultural and linguistic characteristics of the children being assessed.

One of the indicators of effective assessments is that "assessments are designed for and validated for use with children whose . . . cultures [and] home languages . . . are similar to those of the children with whom the assessments will be used" (NAEYC & NAECS/SDE 2003, 2). In other words, assessments should be culturally and linguistically responsive and appropriate.

INDICATORS

2a. All screenings and assessments used with young English-language learners are culturally appropriate.

Culturally appropriate or culturally responsive assessments are those that occur in settings that embrace diversity and demonstrate esteem for a child's home culture; are administered by bicultural professionals who are knowledgeable about the values and norms (especially norms pertaining to interactions) of the child's home culture; do not include inappropriate referents to objects or words that are either unfamiliar to the child or may carry a different meaning than the one intended; and are interpreted in the context of the child's cultural and social history.

As defined in the full NAEYC and NAECS/SDE (2003) position statement, the term culture includes ethnicity, racial identity, economic class, family structure, language, and religious and political beliefs. Each of these aspects of a child's identity, heritage, and experience profoundly influence the child's development and relationship with the world. The issue of culture is relevant not only to English-language learners, but also to speakers of English who have a unique cultural heritage and may use dialects that differ from those of the prevalent U.S. culture.

Every child deserves learning and assessment environments that are welcoming and responsive to her or his culture (NAEYC 1995). Programs should create environments that respect diversity of language and culture and incorporate elements of children's home language and culture in the classroom's physical environment and activities. Teachers should encourage children to share or explain family values and traditions and to communicate in their home language as well as English. Teachers who speak a child's home language should use it, as well as English, to communicate with the child. A classroom climate that shows that teachers and children value children's cultures and home languages tends to reduce children's sense of intimidation and inhibition and encourage their attempts to communicate (NAEYC 1995). This type of environment is important because it allows teachers more opportunities to observe children's abilities, reduces the chances that a teacher will prematurely or incorrectly conclude that language errors indicate disability, and allows children to show all of their skills and capabilities-leading to accurate assessment conclusions.

The adults involved in conducting and interpreting assessments-they are usually teachers but may be aides or specialists (see section 5)-must be aware of how cultural values may affect young children's behavior and performance on assessments (Soto 1991). They should make a point of knowing generally about a child's culture, such as the important holidays, unique customs and traditions, major figures of that culture, and so forth. However, general knowledge is not enough; those assessing should find out as much as possible about the child's community-for example, adaptations the community has made so as to continue traditions from the country of origin, and specific cultural concerns with which the community may be dealing. This and other information allows those assessing to individualize an assessment to make sure it is culturally appropriate-that is, compatible with the child's interaction and communication style (Bruns & Corso 2001; Santos & Reese 1999).

Culturally shaped expectations and values affect young children's ideas about interactive behaviors, such as when they are supposed to talk, to whom they should talk, and what kind of language to use in various contexts (Espinosa, in press). These factors affect performance during assessments, especially standardized formal assessments in which a child may not know the person conducting the assessment. For example, children from some cultures may be reluctant to use elaborate language when speaking to adults, having been taught to use it only in answer to particular questions or to use formulaic responses. Although they are capable of providing a more sophisticated answer, these children may shrug or give only a short phrase in response to a question, as is appropriate in their culture (AERA, APA, & NCME 1999). An individual familiar with their cultural norms will understand this phenomenon and interpret their responses accordingly.

Before being accepted as culturally appropriate, an assessment should be carefully examined by bilingual, bicultural professionals familiar with the culture and community in question to ensure the assessment is culturally appropriate. Culturally appropriate assessments do not contain any inappropriate referents, such as words and objects that would be unfamiliar or have an unintended meaning for a child. An example of an inappropriate referent is a picture of Raggedy Ann, with which a child may not be familiar because it is unique to the prevalent U.S. culture (Santos 2004). Another example is a picture of a bear. Teddy bears may be meant to represent

cute, benign, friendly animals; but in the Navajo culture, bears usually denote something wicked (Nissani 1993). Differences in connotation like this can result in confusion, frustration, and misunderstood responses on the part of the child. Assessments, whether standardized formal assessments or classroom-based observation tools, should avoid culturally inappropriate components.

If the individual conducting an assessment is not familiar with a child's culture, a cultural guide (a qualified representative of the child's cultural and linguistic group who can serve as a broker or mediator) should assist in the assessment process, including the interpretation of results. The presence of a person who knows the child's culture helps ensure that assessment methods and measures are appropriate, and that the child can communicate in a language, dialect, or interaction style that is comfortable. A cultural guide also should ensure that neither translation discrepancies nor cultural conventions nor differences in childrearing practices lead to misinterpreted results (Santos 2004).

Interpretations of assessment results, whether from systematic observations or direct assessment, should be made only in the context of a child's language history and cultural background. Everyone involved in the assessment process must consider the child's culture, home context, social history, and prior experiences and learning opportunities before drawing conclusions about the child's performance (or the performance of a subgroup) on assessment procedures and before making decisions that will affect the child's education and receipt of services.

2b. All screenings and assessments used with young English-language learners are linguistically appropriate.

An assessment approach that is linguistically appropriate or responsive goes beyond simply translating materials into another language. Linguistically appropriate assessment takes into account a child's language history, proficiency, and dominance and preference, where applicable; has alignment between the goal of the assessment and the language(s) used to assess; is administered by a bilingual person fluent in the language of the assessment; and allows for flexibility in the child's language of response (except when assessing for proficiency in a given language). Because of these challenges, when assessing young English-language learners, it is important to include curriculum-embedded, observational assessments and other methods that place less reliance on children's production or comprehension of language as a key part of the assessment. However, to some degree all assessments are measures of language (AERA, APA, & NCME 1999; Shepard, Kagan, & Wurtz 1998), and the issues analyzed below are important to keep in mind no matter what the assessment purpose or approach.

Language history and proficiency. Whether assessments are classroom based or part of a larger assessment system, planning for assessment of young English-language learners should begin with gathering information about the child's and family's history with language. The information should include the language the family primarily speaks at home and in the community, other languages spoken in the home, the family's country of origin, the length of time the family has lived in the United States, the child's age at first exposure to English, and who in the family speaks English and how well (CLAS Early Childhood Research Institute 2000; Santos & Reese

1999). Program staff is often able to do this in preliminary home visits or family meetings. (See section 5 for more about family involvement in assessment.)

Also, for children relatively advanced in their language development, those assessing need to determine a child's language dominance (the language in which the child is most proficient) and language preference (the language in which the child prefers to speak) (Páez 2004), keeping in mind that these characteristics are difficult to determine with very young children whose language development is rapid, variable, and dependent on the home-language environment. Accurate assessment of language proficiency is especially important for young English-language learners, because they may seem to be speaking English with ease when actually they are not fully capable of understanding or expressing themselves in complex ways and still lack vocabulary skills, auditory memory, ability to follow sequenced directions, and other markers of proficiency (NAEYC 1995). Insights about language proficiency will help teachers and others more effectively plan learning opportunities for young English-language learners.

Cautions about language proficiency assessments. Assessments of language proficiency should rely only on instruments and procedures designed to assess language proficiency, not those designed to assess content knowledge or anything else. It should also be noted that some researchers have concerns about the validity of English proficiency tests. For example, there is little evidence that content and construct validity of English proficiency tests align sufficiently with the standards put forth by experts in the field, such as Teachers of English as a Second Language (Bailey & Butler 2003).

Furthermore, many language proficiency assessments are not consistent in how they measure various aspects of language and measure only a limited set of language components. For example, one assessment evaluating oral language proficiency might measure ability to follow instructions (a component of basic interpersonal communication skills), whereas another might measure knowledge of synonyms and antonyms (a cognitive-academic language proficiency). It is important not to assume that all assessments of language proficiency measure the same aspects of language (Schrank, Fletcher, & Guajardo Alvarado 1996). These cautions are not meant to deter assessment of language proficiency, but rather to prompt assessment decision makers to carefully review information about language proficiency assessments before selection.

Home language or English? Matching the method and purpose of assessment. After gathering information about the child's language history and current language proficiency, administrators and others responsible for assessment need to consider the purpose of the assessment before deciding on appropriate language(s).

If an assessment is to be used for program evaluation or accountability purposes, it should take place in the language and dialect in which the child can best show what he or she knows and can do. If the child is proficient in both the home language and English and it is unclear which language is dominant, the child should be assessed in both languages. Code-switching (see later discussion) should be allowed. Although it is always important that a well-trained professional fluent in the child's home language and knowledgeable about the child's home culture administer any assessments to be used for these purposes, it is especially important here.

If an assessment is to be used to guide instruction and for other learning-related purposes, three options could be appropriate, depending on the goal of the assessment and the child's level of proficiency: (1) assess only in the child's home language (for example, when evaluating a child's knowledge of content in a specific area, such as mathematics); (2) assess in a language in which the child is proficient, even if it is not the child's home language (this could be English or a third language); or (3) assess in both English and the child's home language. Because of the episodic, unpredictable, and rapidly evolving nature of language development among young English-language learners, a dual-language approach is recommended, assessing in both English and the child's home language whenever possible.

The dual-language approach. There are several reasons to recommend a dual-language assessment approach. There is no one path to learning a new language (Wong Fillmore 1985); there are multiple environmental influences and individual differences that interact to shape second-language acquisition. To get an accurate picture of progress in the language domain, it is therefore useful to monitor progress in both English and the child's home language. Also, young children perform better when the language of the assessment matches the language of instruction (Abedi, Leong, & Bodrova 2004; Gonzalez, Bauerle, & Felix-Holt 1996). Young children are "instructed" both at home and in the education setting, but the content and the skills developed may be different, and they may be differently revealed in assessments in either the home language or English-but not equally well revealed in both languages. For these and other reasons, whenever feasible, assessment should involve both languages to most accurately reveal children's knowledge and skills.

Code-switching. When learning a second language, children often go through a period of code-switching or code-mixing, using rules and words of both languages from one sentence to another or within a sentence, respectively (Chamberlin & Medinos-Landurand 1991). (Even as adults, bilingual individuals often mix languages in social conversation with others of the same group.) This behavior is not unusual and is not necessarily a sign of deficiency in language development (Garcia 1990). It demonstrates children's efforts not only to practice multiple languages, but also to successfully navigate multiple cultural markers, norms, and values in order to communicate effectively (Celious & Oyserman 2001). Except when evaluating language proficiency in a particular language, those conducting assessments should accept responses that involve children's code-switching and code-mixing as an appropriate means of determining what children know and can do.

2c. Translations of English-language instruments are carefully reviewed for linguistic and cultural appropriateness by native speakers well versed in the complex issues of assessment and translation.

Assessments used with English-language learners are often translations of assessments developed for monolingual English-speaking children. There are a number of things to consider when selecting and using translated materials (see Santos et al. 2001), and everyone who is involved in assessment of young English-language learners should be aware of these considerations. For example, it is common to assume that a translated assessment is appropriate for a young English-language learner simply because the language of the assessment is the child's home language. This assumption may not be correct. Translated materials are likely to

differ from the original version in both content and construct, and those conducting the assessment should not assume a translation produces a version of the instrument that is equivalent to the original version in difficulty, content, and reliability and validity (AERA, APA, & NCME 1999; Kopriva 2000). Translations may use a dialect, colloquialisms, and unfamiliar referents that are inappropriate for the child being assessed. Spanish-translated materials appropriate for a child from a Mexican American community, for example, may not be appropriate for a child from a Puerto Rican community.

Native speakers of a child's home language who are familiar with assessment constructs should carefully review translated materials for cultural and linguistic appropriateness (Ohtake, Santos, & Fowler 2005; Santos & Reese 1999). Likewise, test developers should establish translation equivalence-evidence that the adapted instrument is comparable to the original in content and difficulty-before assessment decision makers decide to use translated instruments (AERA, APA, & NCME 1999). Methods of checking for appropriateness and equivalence can include back-translating the translated version of the assessment back into English to determine whether the home-language and English-language versions are the same in terms of content and difficulty. However, back-translation alone is not sufficient; assessments still need to be examined for reliability and validity. On-the-spot translations of standardized assessments should not be used (Páez 2004), as they are likely to include errors and are highly unlikely to be appropriate and equivalent at the necessary levels.

3. Characteristics of Assessments Used to Improve Instruction

RECOMMENDATION

The primary purpose of assessing young English-language learners should be to help programs support their learning and development; classroom-based assessment should maximize the value of the results for teachers' curriculum planning and teaching strategies.

The indicators discussed in this section are adapted from those outlined in the full position statement on curriculum, assessment, and program evaluation (NAEYC & NAECS/SDE 2003), but with special reference to implications for young English-language learners.

INDICATORS

3a. Programs rely on systematic observational assessments, using culturally and linguistically appropriate tools as the primary source of guidance to inform instruction and to improve outcomes for young English-language learners.

For monitoring progress, informing classroom teaching, and improving a child's learning, assessments based on observation provide the richest and most relevant, accurate, and useful data. However, observation needs to go beyond incidental or casual processes; observation-based assessments should be chosen with care to ensure they are sound, of high quality, and culturally and linguistically appropriate. When used systematically as part of an assessment system, they should have appropriate evidence of reliability and validity. Program staff involved in teaching young children should rely most heavily, therefore, on observational assessments such as rating

scales, checklists, analyses of samples of children's work, and portfolio approaches, many of which are linked to a particular curriculum model. These methods are especially valuable in assessing young English-language learners, whose strengths and developmental needs may not reveal themselves through direct verbal methods.

Observational assessments alone, however, are not sufficient for all purposes. In some cases, standardized formal assessments may be not only useful but also necessary—for example, when assessing for certain disabilities. When used as part of a comprehensive assessment system, information from standardized formal assessments also can be helpful in monitoring children's progress. The next section, section 4, discusses the appropriate uses of standardized formal assessments with young English-language learners.

3b. Assessments for young English-language learners are based on multiple methods and measures.

No one assessment, measure, or method of collecting information about a child will provide all the information educators and others want to know. This is especially true for young English-language learners, and assessments of any aspect of their development and learning should always include several methods and measures (Gonzales, Bauerle, & Felix-Holt 1996). Because purely verbal procedures tend to underestimate children's cognitive ability (Chapman 1991; Gonzalez 1994), approaches should include both verbal and nonverbal procedures. Also, as with all young children, assessments should occur across all the domains of the curriculum and should involve a range of activities. Allowing children—especially young English-language learners—to express themselves and to be assessed across the curriculum in areas as diverse as art, music, and block building gives them opportunities to demonstrate their intellect and knowledge in ways that exceed the boundaries of language (NAEYC 1995). To round out this picture, observations should occur across different settings, such as in the classroom, on the playground, and during interactions with peers, familiar adults, and strangers.

3c. Assessments for young English-language learners are ongoing; special attention is given to repeated assessments of language development over time.

Young children's development and learning can never be captured in a single snapshot. Ongoing assessment is always the recommended practice. Special issues around language learning make this point especially relevant in assessing language development among young English-language learners. Learning a new language takes time. There is a misconception that young children acquire language more easily and quickly than adults; in fact, with the exception of pronunciation, this is not the case (Soto 1991). Children can, but do not necessarily, achieve social language proficiency in a second language in two to three years and academic proficiency in four or more years (NAEYC 1995). Because of the long-term nature of second-language development, and because paths to proficiency are uneven and unpredictable, a snapshot approach to assessment is particularly ineffective for young English-language learners. A more accurate picture of a child's progress will reveal itself gradually over time as a child experiences a variety of social interactions and opportunities for growth in all domains. Complex interactions between children's social, linguistic, and cognitive domains determine what path language development might take, and individual differences among children lead to great variability in

those paths (Genishi 1989; Wong Fillmore 1985). For these reasons, assessments used to monitor and guide children's learning in language and other domains should be ongoing (Duarte & Gutierrez 2004; Santos 2004; Trister Dodge et al. 2004), with emphasis on assessment in everyday, naturalistic settings.

3d. Assessments for young English-language learners involve two or more people.

Conclusions about the development of young English-language learners should always be based on information from multiple sources (Gonzalez, Bauerle, & Felix-Holt 1996; Lewis, 1991; Ohtake, Santos, & Fowler 2005; Páez 2004). Assessments usually involve some interpretation and judgment on the part of those assessing. Because of this subjectivity, there is always room for error and bias in the assessment process (Espinosa, in press). With assessments of young English-language learners, the backgrounds of those assessing—their identity, cultural stereotypes, life experiences within linguistic and cultural milieus, conceptualizations of constructs measured, and so forth—can influence assessment decisions (Gonzalez, Bauerle, & Felix-Holt 1996). (Section 5 offers recommendations concerning the characteristics of those who assess young English-language learners.)

Furthermore, because there are few appropriate instruments for young English-language learners, it is important to triangulate information, or verify it by getting information from a number of people, especially when the results of assessment have important consequences. Adults often have different perceptions of a child's abilities, depending on the sources of information available and the settings in which the child and adult interact. This difference in perception can become particularly salient when there is a linguistic divide—when adults communicate with the child in different languages. Observations or data about a child can more safely be assumed to be accurate if they are verified by several people (such as a teacher, a parent, and a reading specialist) rather than by only one person.

More than one professional (teacher, paraprofessional, consultant, and so forth) should be involved in significant assessment-related decisions about a child's progress, and at least one of these professionals should be proficient in the child's home language. In addition, at least one of the people providing input on the child's progress should be a family member. (See section 6 for discussion on the role of family members.)

3e. Assessments for young English-language learners are age appropriate.

Age is an important consideration in selecting assessment tools and procedures for all young children; assessments used for preschool children should obviously look quite different from assessments used with children in primary school. Because there are few assessments—and in some cases, no assessments—available for young English-language learners that are psychometrically, linguistically, culturally, and age appropriate, those who assess may be tempted to use an assessment designed for an age group different from the age of the child being assessed, if that assessment tool has other positive features. Despite these constraints, assessment decision makers should avoid selecting assessments that are developmentally or age inappropriate, as the results are likely to be inaccurate and uninformative.

4. Using Standardized Formal Assessments

RECOMMENDATION

The development of state and other accountability systems has led to increased use of standardized formal assessments of young children. Specific considerations about the development and interpretation of these assessments should guide their use with young English-language learners.

Standardized formal assessments, or direct assessments, are typically administered at a single point in time, either orally via questions and answers or, for 6- to 8-year-olds, via paper-and-pencil approaches. When used appropriately and in context, these types of assessments can provide important and useful information. However, early childhood professionals should be aware of concerns about the use of many of these assessments with young English-language learners. The decision to use a standardized formal assessment with young English-language learners should be made cautiously and with awareness of the complexity of the issues involved.

INDICATORS

4a. It is appropriate to use standardized formal assessments to identify disabilities or other special needs and for program evaluation and accountability purposes. They may also contribute to monitoring and improving learning at an individual level as part of a more comprehensive approach to the assessment of young English-language learners.

Because program evaluation and accountability assessment procedures necessarily involve large groups of children, these procedures have primarily relied on standardized formal assessments. When tools and practices are developmentally, culturally, and linguistically appropriate, it may be useful to employ standardized formal assessments for these purposes, keeping in mind the cautions outlined in indicators 4b and 4c.

In addition, because information from standardized formal assessments allows staff to compare a child's progress against the progress of similar children, this information may offer an improved understanding of an individual child's development, if the information is viewed in a broader assessment context. For example, the results of a linguistically appropriate standardized formal assessment could reassure teachers and parents that a young English-language learner who seems to be silent much of the time is actually developing typically, considering her age and language experiences. It is important to reemphasize that only meaningful comparisons should be made; data from English-language learners should be compared to data from other, similar groups of English-language learners and not to monolingual English-speaking children.

4b. Decision makers and those conducting assessments are aware of the concerns and cautions associated with using standardized formal assessments with young English-language learners.

Those responsible for making decisions about assessment systems should be aware of specific concerns about using standardized formal assessments with young English-language learners. They should know, for example, that standardized formal assessments often contain a great deal

of material for which comprehension depends on children's previous learning experiences and background knowledge rather than on their cognitive functioning (Kozulin & Garb 2001). True levels of cognitive ability tend therefore to be underestimated for young English-language learners when using standardized tests (Gonzalez 1994; Gonzalez, Brusca-Vega, & Yawkey 1997).

In addition, and as detailed in indicator 4c, there are serious concerns about the validity and norming of standardized formal assessments used with English-language learners (Abedi 2002; Navarrette & Gustke 1996; Zehler et al. 1994). In many cases there is simply no information about the validity of assessments used (Gutierrez-Clellan & Kreiter 2003).

4c. Decision makers and test developers carefully attend to test development issues, including equivalence and norming.

NAEYC urges the rapid development of new and better assessment tools that will allow young English-language learners to be assessed in ways that benefit them. However, a number of key issues need careful attention as these assessments are developed.

Equivalence across versions. Ideally, standardized instruments used with populations of young English-language learners are developed through an iterative or concurrent process in which items originate from both languages of the versions being developed. Equivalence across versions of the instrument being developed must be established at several levels. First, the versions should have construct equivalence, or evidence that what the instrument measures for one child is the same as what a version of it measures for another child (for example, it measures academic knowledge for all children; it does not measure academic knowledge for monolingual English speakers and language proficiency for English-language learners). Second, versions should have functional equivalence, meaning that the activities or behaviors measured have the same meaning in each cultural or linguistic group being assessed. Third, they should have translation equivalence, meaning that if instruments are translations, they are comparable in content to the original. And fourth, they should have metric equivalence, meaning that scores from each version of the instrument have comparable psychometric properties, such as reliability and validity (AERA, APA, & NCME 1999). The linguistic and cultural characteristics of each of the groups of children for which the instrument is intended should be reflected in the samples used throughout the processes of test design, validation, and norming (AERA, APA, & NCME 1999).

Norm-referenced assessments. Norm-referenced assessments are standardized so that a child's performance or score is interpreted in relation to the performance of a group of peers who have previously taken the same test. However, these assessments only lead to useful insight when the instruments and standards have been appropriately developed and when the comparisons would make sense. Norms for assessments to be used with young English-language learners should be based on the performance of other young English-language learners rather than on the performance on monolingual children-including children monolingual in the child's home language (Mazzeo et al. 2000; Navarrett & Gustke 1996; Zehler et al. 1994). Moreover, norms should be based on similar populations of children. If, for example, a Spanish-language version of an assessment will be used with Mexican, Puerto Rican, Cuban, Central American, and

Spanish children, then norms, reliability, and validity should be established with members of each of these groups (AERA, APA, & NCME 1999).

At present, few assessments are normed this way. Until more appropriately normed assessment tools are available, those selecting assessments and using their results with young English-language learners should pay close attention to how the tools were normed and exercise caution when interpreting the results.

4d. Decision makers and those conducting assessments know appropriate conditions for using and interpreting standardized formal assessments with young English-language learners.

Given the concerns, decision makers and teachers should know which assessments might be appropriate to use with young English-language learners. As already emphasized, appropriate standardized formal assessments are those that (a) meet the highest psychometric or technical standards, showing clear evidence of validity and reliability; (b) are used only for the purpose for which the assessment was designed; and (c) are based on norms from similar populations of young English-language learners. Again, few assessments fully meet these requirements; assessment decision makers should therefore exercise caution in how they use information from assessments with respect to young English-language learners.

If standardized formal assessments are used with young English-language learners, it may be appropriate to incorporate accommodations to allow children to show a true picture of their abilities. It may be appropriate, for example, to allow greater wait time for some items, rephrase directions and questions so a child can understand them, and ask for explanations to clarify the child's thinking. Those assessing should plan for additional time in the assessment process to (a) assess language proficiency before selecting measures to assess knowledge and abilities; (b) obtain background information about the child; and (c) conduct additional procedures that might be necessary (Páez 2004).

In addition, standardized formal assessments should emphasize children's progress over time, as other assessments do, and results generally should be interpreted in the context of children's progress or growth rather than on an absolute basis.

5. Characteristics of Those Conducting Assessments

RECOMMENDATION

Whatever the purpose of the assessment, those conducting assessments of young English-language learners should have cultural and linguistic competence, knowledge of the children being assessed, and specific assessment-related knowledge and skills.

Even the most linguistically and culturally appropriate assessments may be inappropriate and ineffective if the adults who are implementing the assessments and interpreting their results lack

relevant experience and preparation. This section explains who should be responsible for assessing young English-language learners and what those adults should know and be able to do.

INDICATORS

5a. It is primarily teachers who assess young English-language learners, but paraprofessionals, assessment assistants, and specialized consultants also play an important role.

Depending on the purpose of an assessment, a variety of individuals may conduct and interpret the results of the assessment. Because the primary purpose of early childhood assessment is to help teachers learn more about children in order to make informed classroom-level decisions about curriculum and teaching practices, most often those involved in assessing are-and should be-children's teachers. In high-quality early childhood programs, teachers assess children on a daily basis, using systematic, well-validated observational methods, analysis of children's work, and other assessment approaches that are developmentally, culturally, and linguistically appropriate.

Despite the primary role of the classroom teacher, paraprofessionals (e.g., teacher aides), assessment assistants from the community, and specialized professional consultants also play an essential role in some aspects of the assessment of young English-language learners. Programs should be proactive in establishing a pool of assessment assistants on whom they can call as need arises (Páez 2004). Depending on the roles for which they are needed, this pool might include community leaders, business leaders, and members of the clergy who are from the child's cultural and language community (Bruns & Corso 2001). Before collaboration, programs should determine potential assistants' personal history related to the target language and culture as well as other qualifications, including written and oral language proficiency (Páez 2004).

Besides directly helping with some types of assessments, these assistants, who should be fluent in the child's home language (and English) and familiar with the child's community, may be excellent resources to serve as cultural guides or cultural-linguistic mediators between home and school (Lynch & Hanson 2005). They can facilitate communication and understanding between program staff and families and can teach staff unfamiliar with a child's culture about appropriate ways to interact with family members and about community beliefs and values (Dennis & Giangreco 1996; Kalyanpur & Harry 1999). Depending on their qualifications and skills, they also can serve essential roles in translating materials and reviewing already translated materials and in interpreting before, during, and after assessments-especially when standardized formal assessments are being conducted. However, they should not be solely responsible for administering assessments unless they are qualified and have been trained specifically to do so.

Finally, adults with specialized professional training also play a part in assessment of young English-language learners-for example, when screening results indicate the need for in-depth diagnostic assessment or when certain assessments are externally administered as part of an accountability system. These individuals too require knowledge relevant to the assessment of young English-language learners, and they require the ability to conduct assessments in the

child's home language as needed. In some situations, community assessment assistants may serve as helpful partners in this effort.

Whatever their role, it is important that teachers, paraprofessionals, and consultants who are involved in any aspect of the assessment process know the relevant laws and ethical issues, the purpose of various assessments, and the importance of using correct procedures.

5b. Those assessing young English-language learners are bilingual and bicultural.

Ideally, those assessing should not only be fluent in the child's home language but also be familiar with the dialect spoken in the child's community. Those who assess young English-language learners must appreciate diversity and show respect for the dignity and uniqueness of all people. People who hold prejudices or negative stereotypes about groups of children based on their background should not assess young English-language learners. Teachers and others assessing should know the cultural traditions, values, and beliefs of the children they assess and should be aware of generally preferred interaction styles for people from those cultures. They should know not only about the child's culture generally but also about the child's current community specifically—its goals, challenges, and unique circumstances.

5c. Those assessing young English-language learners know the child.

Children tend to perform better when they know and feel comfortable with the person assessing them (Gonzalez, Bauerle, & Felix-Holt 1996). The person assessing should be someone with whom the child is familiar and comfortable. In the case of ongoing assessment that informs curriculum and instruction, the teacher conducts the assessments and should therefore be familiar with the children who are being assessed. Early in the year, or especially in the case of children whose language and culture differ from that of the teacher, teachers may need additional time and support to build relationships that will contribute to more effective assessment and interpretation of assessment results. In situations where an external adult administers standardized formal assessments, that individual should spend time with and develop rapport with the child before the assessment.

5d. Those assessing young English-language learners are knowledgeable about language acquisition, including second-language acquisition.

Whether they are conducting language assessments or assessments in other domains, teachers and other professionals assessing young English-language learners should know about the development of language proficiency and specifically about second-language acquisition, both sequential and simultaneous. Too often, children from diverse backgrounds are overrepresented in special education programs, so it is important for those assessing to be aware that language errors as a function of learning stage might incorrectly lead to diagnosis of a disorder or developmental disability (Espinosa, in press). For example, an untrained teacher might mistake low language assessment scores for a reading disability, when in fact the child is simply not proficient in English, the language of the assessment. In the beginning stages of a child's second-language acquisition, it can be difficult for the individual assessing to separate mere learning errors from disability or delay, so he or she should be aware of ways in which the behaviors look

similar in order to reduce the frequency of incorrect conclusions. Also, those assessing should know which specialists—including English as a Second Language teachers, speech and language pathologists, and reading specialists—to consult for assistance.

5e. Those assessing young English-language learners are trained in and knowledgeable about assessment in general and about considerations in the assessment of young English-language learners in particular.

As emphasized in NAEYC's standards for early childhood professional preparation, well-prepared early childhood professionals understand the goals, benefits, and uses of assessment, and they practice responsible, ethically grounded assessment (Hyson 2003; NAEYC 2005). In addition to knowing about assessment in general, those being prepared to work with young children should be trained in and knowledgeable about the assessment of young English-language learners, including knowing about specific ethical issues that may arise. They should also know about selecting appropriate assessments, soliciting information from family members, consulting with cultural guides, using translators, interpreting results, and the purposes for which assessments may be used.

6. The Role of Family in the Assessment of Young English-Language Learners

RECOMMENDATION

Families of young English-language learners should play critical roles in the assessment process, being closely involved in a variety of appropriate ways.

Families are always significant sources of information about their young children, and in the case of young children with disabilities, there are legal requirements for family involvement. Efforts to gather information and build positive relationships with families whose home language is not English are essential for many reasons, one of which is to create effective assessments that will benefit children. Yet family members should not be exceptionally burdened or asked to take on roles for which they are neither prepared nor responsible.

INDICATORS

6a. Professionals involved in the assessment of young English-language learners seek information and insight from family members in selecting, conducting, and interpreting assessments.

Family members have perspectives on, preferences for, and observations about the child that program staff will not know unless they ask—and they should ask. Program staff should seek this critical information from parents, grandparents, and other caregivers in the home, listening respectfully and with an open mind to the family's goals and concerns for the child, as well as what behaviors and skills the family observes in the child (Banks, Santos, & Roof 2003; Santos, Corso, & Fowler 2005). Important for the families of all young children, these discussions are especially valuable when families are linguistically and culturally diverse.

However, cross-cultural differences too often interfere. Lack of experience with diverse families often disrupts the process of developing positive, respectful relationships between those assessing and family members, resulting in lack of family input (McLean 2002). Even if the individual assessing is generally familiar with a culture, there are significant within-culture differences and within-family differences about which teachers and others involved in assessing should become aware.

If program staff unwittingly offend families, families are unlikely to be forthcoming with important assessment-related information (Dennis & Giangreco 1996), so one of the first things those assessing should determine is a family's preferred communication style. For example, they should determine whether a family prefers an informal, friendly relationship with program staff, as do many traditional Hispanic families (Gonzalez-Alvarez 1998), or whether the family prefers a more formal, professional relationship with program staff, as do some traditional Asian families (Schwartz 1995). Each family is unique, so although sensitivity to general cultural differences is an important foundation for good communication, those assessing need to learn about the characteristics and preferences of individual families.

When comfortable patterns of communication between families and staff have been established, individuals conducting assessments (in this case, classroom teachers or other program staff) should gather as much information as possible about each family's history and current situation, as these factors could affect a child's responses to both observational and direct assessments and should affect interpretation of results. Those assessing should keep in mind that many factors work together to influence family functioning, so looking at a factor in isolation may be misleading.

Teachers and others involved in assessing young English-language learners should know each family's country of origin, where a family currently lives, how long the family has lived in the United States, and the primary language the family speaks at home and in the community. It is especially important to determine whether a family has any concerns about their child's language development. They also will want to find out, in a sensitive manner, about a family's education, religious affiliation, and degree of acculturation. If possible, teachers and others assessing should sensitively seek additional information, such as whether there are specific accomplishments a family is proud of, what the family believes are the most important things children should learn, and how the parents see their role (Santos & Reese 1999). They should determine families' concerns about stereotyping, prejudice, and discrimination. They should also, if possible, find out about families' experiences with educational, health, and other institutions. These concerns and experiences are likely to affect families' involvement and interactions with their child's educational program and teachers, as well as their willingness to share assessment-related information.

6b. Programs refrain from using family members to conduct formal assessments, interpret during formal assessments, or draw assessment conclusions.

There are some types of assessment practices in which it is appropriate and important to involve family members, for example, in completing observation charts used to record behaviors in the

home, in completing parent rating scales or questionnaires, and in dynamic assessments that involve observations of parent-child interactions.

Family members, however, are not trained in administering formal or standardized assessment instruments, and they also are less likely than professionals to be objective about the performance of their child. Aside from the types of circumstances just noted, family members should not conduct assessments of their child, nor should they serve as interpreters during formal assessments of their child. Regardless of whether formal or informal assessment procedures are being used, family members should not be responsible for independently interpreting assessment results or drawing conclusions from the assessments, although their insights are an essential contribution to the interpretation. Because of confidentiality issues and the sensitive nature of assessment results, it is also recommended that close friends of the child's family not be involved in assessment procedures.

Program staff may find themselves in situations where the only person available who speaks both English and the child's home language is a family member, often an older sibling (or another older child in the community). Children of any age may be linguistically and cognitively ill-equipped to participate, even as interpreters, in assessment procedures; they also may be burdened emotionally by participating in an assessment of their younger sibling. Instead of relying on parents, siblings, or older children in the community, those assessing should make every effort to collaborate with a professional consultant or an assessment assistant who is familiar with a child's community but not intimately involved with the child's family. If it is simply not feasible to collaborate with a professional consultant or cultural guide, decisions about involving older children should be made cautiously on a case-by-case basis.

6c. Professionals involved in assessment regularly inform and update families on their child's assessment results in a way that is easily understood and meaningful.

Staff and family communication about a child's development is a two-way street. Program staff, who should be soliciting assessment-related information from family members on a regular basis, should also be giving information regularly to family members about results of assessments. Regardless of what role family members play in assessment procedures-whether they are intimately or only remotely involved-staff should explain what the results show about the child's development in a way that is easily understood and meaningful to the family. At the same time, staff should seek to understand families' interpretations of assessment results and how the results may fit with families' goals for their children-in language acquisition and in other areas. Clearly communicating results and providing guidance about how to use the information are essential components of responsible assessment for staff working with all families; but this component is especially important for staff working with families from diverse cultures and who might have language differences. Staff should be particularly proactive in finding clear, meaningful ways to share assessment information reciprocally and respectfully with families of young English-language learners.

7. Needs in the Field

RECOMMENDATION

Resources should be invested to ensure rapid progress on several fronts: expanding the knowledge base; developing more and better assessments; increasing the number of bilingual and bicultural professionals; and creating professional development opportunities for administrators, supervisors, practitioners, and other stakeholders in effective assessment of young English-language learners.

The supports and resources available to those developing assessment-related policies, designing assessment tools and procedures, and directly assessing or supervising those who assess young English-language learners have not kept pace with rapidly increasing demands in the field. For the preceding recommendations to be implemented, specific actions are needed.

INDICATORS

7a. Scholars provide an expanded knowledge base about second-language acquisition and the development of young English-language learners.

Program administrators, teachers and other program staff, psychologists, and other professionals and paraprofessionals who work with young English-language learners need practical information about second-language acquisition. They need to know how second-language acquisition relates to cognitive, emotional, cultural, and social factors and how to monitor it effectively. They also need to know more about what influences the development of young English-language learners, especially about the factors that may be under their control. Researchers must help the field move forward with more fully developed theoretical frameworks and empirical research about second-language acquisition as it pertains to young children in general, and specifically how it relates to children from various language and cultural groups. Continued efforts to develop an expanded knowledge base in this area must be supported as an essential foundation for evidence-based assessment policies and practices.

7b. More and better assessments are developed to meet the most pressing needs.

Those responsible for planning and conducting assessments of young English-language learners have few, and sometimes no, appropriate assessments from which to choose. Because 79 percent of English-language learners in U.S. public schools are Spanish speakers (Abedi, Hofstetter, & Lord 2004), the first priority should be to develop appropriate assessments in Spanish. In developing these assessments, assessment designers should be responsive to the within-group differences in dialect and culture that will be represented among Spanish-speaking children. Beyond these needs, research and technical expertise are needed to develop and improve assessments for young children who speak the many other languages represented in U.S. early childhood programs.

7c. Policy makers, institutions of higher education, and programs adopt policies and practices to recruit and retain a diverse early childhood workforce, with a focus on increasing the number of bilingual and bicultural early childhood professionals.

The field urgently needs culturally sensitive bilingual early childhood professionals who not only are proficient in conducting assessments but also can communicate with children and family members in their home language. As the demographics of the United States shift to include greater and greater numbers of bilingual and multilingual children, the need for the early childhood workforce to diversify becomes more urgent. So far, the workforce has not kept pace with the diversity of the children served (Lynch & Hanson 2005).

Specifically, the field needs to increase the number of well-prepared bilingual professionals. Bilingual teachers will be able to support the preservation of young children's home language and culture. They help create environments that encourage young English-language learners to participate in social interactions, and they are likely to empathize with children's challenges, frustrations, and ultimate satisfaction as they attempt to learn a new language. Teachers who are not able to become fully bilingual benefit from learning even the basics of a second language. Besides enhancing communication with children and families, experiencing the process of learning a second language may help professionals be more sensitive to the challenges and processes experienced by young children learning a new language.

7d. Early childhood professionals, including program administrators, receive ongoing opportunities for professional development and support in the area of assessing young English-language learners.

To improve the quality of assessment practices with young English-language learners, the early childhood field needs teachers and administrators who know about assessment principles and practices, how young children acquiring a second language typically develop, and the implications of second-language acquisition for assessment. Many early childhood teachers receive little preparation for working with children and families from a wide range of cultures and linguistic backgrounds (Garcia et al. 1995). In-service and preservice educators may not be giving enough attention to working with young English-language learners; for example, only 10 percent of baccalaureate and 8 percent of associate degree early childhood programs require a course on working with young English-language learners (Early & Winton 2001). And simply managing day-to-day responsibilities with limited resources and time makes it difficult for program staff to attend to this need.

The burden, however, should not rest primarily on teachers. Assessments, especially standardized formal assessments, are usually selected and implemented by education decision makers as part of program-level assessment planning or large-scale accountability systems. Program administrators should be actively involved in implementing and acting on results of assessments to improve outcomes for children and their programs. For example, they should select, or help staff select, good assessments for young English-language learners; they should monitor assessment procedures; and they should assist teachers in using results to inform their curriculum and teaching practices. They also need to be informed about the development of young English-language learners in order to create a program philosophy and environment that reflects a coherent, knowledge-based, mission-driven approach to the assessment and teaching of young English-language learners.

Given the growing urgency and the demands related to assessment of second-language learners, teacher educators, program administrators, trainers, and policy makers need to make this area a priority, investing resources and emphasizing its importance to staff, students, and the public. Professional development opportunities in the form of workshops, conference sessions, college courses, and distance-learning activities should be developed, widely accessible, and linked to incentives for participation.

CONCLUSIONS

If well implemented, the recommendations presented in this document would contribute to more positive developmental and educational outcomes for the millions of young English-language learners served by early childhood programs. At present and as emphasized throughout this document, the conditions needed to fully implement the recommendations do not yet exist, although promising practices are evident in many settings-practices that need to be better identified and showcased as models. If the vision behind these recommendations were fully realized, then technically sound and developmentally, linguistically, and culturally appropriate assessments would be available for all purposes and settings. The recommendations also reflect a vision in which all early childhood program managers and practitioners would be fully prepared to assess the diverse children in their programs in ways that support those children's learning and development. And the recommendations envision a policy environment with both the resources and political will to support the needs of young English-language learners and their families. At present, sufficient resources are not being directed toward these ends.

Until more resources and supports-financial, scientific, and professional-are made available, early childhood professionals will have to continue to use their best judgment, wisdom, and practical knowledge to make decisions about how to effectively assess and use assessment results for each child in their care, with the limited means currently available to them. At the same time, early childhood professionals who guide programs and work directly with the millions of young English-language learners must continue to advocate for the support and resources they need in order to fully implement these recommendations.

References

- Abedi, J. 2002. Standardized achievement tests and English-language learners: Psychometrics issues. *Educational assessment* 8 (3): 231–57.
- Abedi, J. 2004. The validity of the classification system for students with limited English proficiency: A criterion-related approach. Manuscript.
- Abedi, J., C.H. Hofstetter, & C. Lord. 2004. Assessment accommodations for English-language learners: Implications for policy-based empirical research. *Review of Educational Research* 74 (1): 1–28.
- AERA (American Educational Research Association), APA (American Psychological Association), & NCME (National Council on Measurement in Education). 1999. *Testing and assessment: The standards for educational and psychological testing*. Washington, DC: AERA. Online: www.apa.org/science/standards.html.

- Bailey, A.L., & F.A. Butler. 2003. *An evidentiary framework for operationalizing academic language for broad application to K–12 education: A design document* (CSE Tech. Rep. No. 611). >Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.
- Banks, R., R.M. Santos, & V. Roof. 2003. Discovering family concerns, priorities, and resources: Sensitive family information gathering. *Young Exceptional Children* 6 (2): 11–19.
- Bruns, D.A., & R.M. Corso. 2001. Working with culturally and linguistically diverse families. Champaign, IL: ERIC Clearinghouse on Elementary and Early Childhood Education. EDO-PS-01-4.
- Celious, A., & D. Oyserman. 2001. Race from the inside: An emerging heterogeneous race model. *Journal of Social Issues* 57: 149–65.
- Chapman, M. 1991. The epistemic triangle: Operative and communicative components of cognitive competence. In *Criteria for competence: Controversies in the conceptualization and assessment of children's abilities*, eds. M. Chandler & M. Chapman, 209–28. Hillsdale, NJ: Erlbaum.
- Chamberlain, P., & P. Medinos-Landurand. 1991. Practical considerations for the assessment of LEP students with special needs. In *Limiting bias in the assessment of bilingual students*, eds. E.V. Hamayan & J.S. Damico, 111–56. Austin, TX: Pro-Ed.
- CLAS (Culturally and Linguistically Appropriate Services) Early Childhood Research Institute. 2000. *Review guidelines for material selection: Child assessment*. Online: <http://clas.uiuc.edu/review/RG-ChildAssessment.html>.
- Collins, R., & R. Ribeiro. 2004. Toward an early care and education agenda for Hispanic children. *Early Childhood Research and Practice* 6 (2). Online: <http://ecrp.uiuc.edu/v6n2/collins.html>.
- Dennis, R.E., & M.F. Giangreco. 1996. Creating conversation: Reflections on cultural sensitivity in family interviewing. *Exceptional Children* 63 (1): 103–16. EJ 529 426.
- Duarte, G., & C. Gutierrez. 2004. Best practices in bilingual early childhood classrooms. *NABE News* 27 (4): 4–7.
- Early, D.M., & P.J. Winton. 2001. Preparing the workforce: Early childhood teacher preparation at 2- and 4-year institutions of higher education. *Early Childhood Research Quarterly* 16: 285–306.
- Espinosa, L. In press. Curriculum and assessment considerations for young children from culturally, linguistically, and economically diverse backgrounds. Special issue, *Psychology in the Schools*.
- Garcia, E. 1990. *Bilingualism and the academic performance of Mexican-American children: The evolving debate*. Charleston, WV: ERIC Clearinghouse on Rural Education and Small Schools. ED321963.
- Garcia, E.E., B. McLaughlin, B. Spodek, & O.N. Saracho, eds. 1995. *Yearbook in early childhood education. Vol. 6: Meeting the challenge of linguistic and cultural diversity in early childhood education*. New York: Teachers College Press. ED 393 560.
- Genishi, C. 1989. The worlds of children: “What Maisie Knew.” *Language Arts* 66: 872–82.
- Gonzalez-Alvarez, L.I. (1998). A short course in sensitivity training: Working with Hispanic families of children with disabilities. *Teaching Exceptional Children* 31 (2): 73–77. EJ 571 917.

- Gonzalez, V. 1994. A model of cognitive, cultural, and linguistic variables affecting bilingual Spanish/English children's development of concepts and language. *Hispanic Journal of Behavioral Sciences* 16 (4): 396–421.
- Gonzalez, V., P. Bauerle, & M. Felix-Holt. 1996. Theoretical and practical implications of assessing cognitive and language development in bilingual children with qualitative methods. *The Bilingual Research Journal* 20 (1): 93–131.
- Gonzalez, V., R. Brusca-Vega, & T.D. Yawkey. 1997. *Assessment and instruction of culturally and linguistically diverse students with or at-risk of learning problems: From research to practice*. Needham Heights, MA: Allyn & Bacon.
- Gutierrez-Clellen, V.F., & J. Kreiter. 2003. Understanding child bilingual acquisition using parent and teacher reports. *Applied Psycholinguistics* 24: 267–88.
- Hepburn, K.S. 2004. *Building culturally and linguistically competent services to support young children, their families, and school readiness*. Baltimore, MD: Annie E. Casey Foundation. Online: www.aecf.org/publications/data/cctoolkit.pdf.
- Hyson, M., ed. 2003. *Preparing early childhood professionals: NAEYC's standards for programs*. Washington, DC: NAEYC.
- Kalyanpur, M., & B. Harry. 1999. *Culture in special education: Building reciprocal family-professional relationships*. Baltimore, MD: Brookes.
- Kopriva, R. 2000. *Ensuring accuracy in testing for English-language learners*. Washington, DC: Council of Chief State School Officers.
- Kozulin, A., & E. Garb. 2002. Dynamic assessment of EFL text comprehension of at-risk students. *School Psychology International* 23 (1): 112–27.
- Lewis, J. 1991. Innovative approaches in assessment. In *Assessment and placement of minority students*, eds. R.J. Samuda, S.L. Kong, J. Cummins, J. Pascual-Leone, & J. Lewis, 123–42. Toronto: Hogrefe.
- Lynch, E.W., & M.J. Hanson, eds. 2005. *Developing cross-cultural competence: A guide for working with children and their families*. 3rd ed. Baltimore, MD: Brookes.
- Mazzeo, J., J.E. Carlson, K.E. Voelkl, & A.D. Lutkus. 2000. *Increasing the participation of special needs students in NAEP: A report on 1996 NAEP research activities* (NCES Publication No. 2000-473). Washington, DC: National Center for Education Statistics.
- McAfee, O., D.J. Leong, & E. Bodrova. 2004. *Basics of assessment: A primer for early childhood educators*. Washington, DC: NAEYC.
- McLean, M. 2002. Assessing young children for whom English is a second language. In *Young Exceptional Children Monograph Series*, no. 4. *Assessment: Gathering meaningful information*, 73–82. Longmont, CO: Sopris West.
- NAEYC. 2005. *Code of ethical conduct and statement of commitment*. Position statement. Washington, DC: Author. Online: www.naeyc.org/about/positions/PSETH05.asp.
- NAEYC. 1995. *Responding to linguistic and cultural diversity: Recommendations for effective early childhood education*. Position statement. Washington, DC: Author. Online: www.naeyc.org/about/positions/PSDIV98.asp.
- NAEYC & NAECS/SDE (National Association of Early Childhood Specialists in State Departments of Education). 2003. *Early childhood curriculum, assessment, and program evaluation: Building an effective, accountable system in programs for children birth through age 8*. Joint position statement. Washington, DC: NAEYC. Online: www.naeyc.org/about/positions/pdf/CAPEexpand.pdf.

- Navarrette, C., & C. Gustke. 1996. *A guide to performance assessment for linguistically diverse students*. Albuquerque: New Mexico Highlands University.
- Nissani, H. 1993. *Early childhood programs for language minority students*. Washington, DC: Center for Applied Linguistics. Online: www.cal.org/resources/digest/Nissan01.html.
- Ohtake, Y., R.M. Santos, & S.A. Fowler. 2005. It's a three-way conversation: Families, service providers, and interpreters working together. In *Young Exceptional Children Monograph Series*, no. 6. *Interdisciplinary teaming*, 33–43. Longmont, CO: Sopris West.
- Páez, D. 2004. Culturally competent assessment of English-language learners: Strategies for school personnel. *Helping children at home and school II: Handouts for families and educators*, S7-1–S7-4. Bethesda, MD: National Association of School Psychologists.
- Santos, R.M. 2004. Ensuring culturally and linguistically appropriate assessment of young children. *Young Children* 59 (1): 48–50.
- Santos, R.M., R.M. Corso, & S.A. Fowler, eds. 2005. *Language and communication: Working with linguistically diverse families*. CLAS Collection #3. Longmont, CO: Sopris West.
- Santos, R.M., S. Lee, R. Valdivia, & C. Zhang. 2001. Considerations when selecting and using early childhood materials translated from one language to another language. *Teaching Exceptional Children* 34 (2): 26–31.
- Santos R.M., & D. Reese. 1999. *Selecting culturally and linguistically appropriate materials: Suggestions for service providers*. Champaign, IL: ERIC Clearinghouse on Elementary and Early Childhood Education. EDO-PS-99-6. Online: <http://ceep.crc.uiuc.edu/eearchive/digests/1999/santos99.pdf>.
- Schwartz, W. 1995. *A guide to communicating with Asian American families: For parents/About parents*. New York: ERIC Clearinghouse on Urban Education. ED 396 014.
- Scott-Little, C., S.L. Kagan, & R.M. Clifford, eds. 2003. *Assessing the state of state assessments: Perspectives on assessing young children*. Greensboro, NC: SERVE.
- Schrank, F.A., T.V. Fletcher, & C. Guajardo Alvarado. 1996. Comparative validity of three English oral language proficiency tests. *The Bilingual Research Journal* 20 (1): 55–68. Online: www.ncela.gwu.edu/pubs/nabe/brj/v20/20_1_schrank.pdf.
- Shepard, L., S.L. Kagan, & L. Wurtz, eds. 1998. *Principles and recommendations for early childhood assessments*. Goal 1 Early Childhood Assessments Resource Group. Washington, DC: National Education Goals Panel.
- Soto, L.D. 1991. Understanding bilingual/bicultural young children. *Young Children* 46 (2):30–36.
- Trister Dodge, D., C. Heroman, J. Charles, & J. Maiorca. 2004. Beyond outcomes: How ongoing assessment supports children's learning and leads to meaningful curriculum. *Young Children* 59 (1): 20–28.
- Wong Fillmore, L. 1985. *Second language learning in children: A proposed model*. Proceedings of a conference on issues in English language development, Arlington, Va. ERIC Document 273149.
- Zehler, A.M., P.J. Hopstock, H.L. Fleischman, & C. Greniuk. 1994. *An examination of assessment of limited English proficient students*. Arlington, VA: Development Associates, Special Issues Analysis Center.

Notes

¹ Although all young children are language learners, we use the term English-language learner to describe young children whose home language is not English, because this is the term used in research and in public policy to describe children learning English as a second language. Many of the issues discussed in this document are relevant for children learning a second language other than English. They are also relevant to trilingual or multilingual young children.

² Some readers of drafts of this publication requested a list of assessments appropriate for use with English-language learners. NAEYC's policy is to refrain from endorsing or recommending specific products; therefore this document does not include such a list.

³ It is important to be aware of federal, state, and local laws, regulations, and rules as well as case law guiding the provision of education, including for immigrant children.

⁴ Assessment procedures for accountability purposes—because they are not designed or used to guide instruction or improve programs—do not directly benefit young children, and the results should never be the sole determinant of any decision made for an individual child, whatever the child's language, culture, or other characteristics.

Copyright © 2005 by the National Association for the Education of Young Children. All rights reserved.

The NAEYC Governing Board, responding to a request from a group of experts in the field, developed this supplement with the assistance of a work group, liaisons, and NAEYC staff. Feedback from additional experts and the public was sought and incorporated. Endorsement is being sought from the National Association of Early Childhood Specialists in State Departments of Education (NAECS/SDE), which the 2003 joint position statement on early childhood curriculum, assessment, and program evaluation, and others.

Work Group

Dick Clifford, *Frank Porter Graham Child Development Institute*

Linda Espinosa, *University of Missouri–Columbia*

Cristina Gillanders, *Frank Porter Graham Child Development Institute*

Virginia Gonzalez, *University of Cincinnati*

Jane Henderson, *California Children and Families Commission*

Luis Hernandez, *Western Kentucky University*

Graciela Italiano-Thomas, *Centro de la Familia de Utah*

Sharon Lynn Kagan, *Columbia University*

Samuel Meisels, *Erikson Institute for Advanced Study in Child Development*

Kellie Rolstad, *Arizona State University*

Catherine Scott-Little, *University of North Carolina–Greensboro*

Patton Tabors, *Harvard Graduate School of Education*

Liaison, NAECS/SDE

Rooney French, Kentucky Department of Education

Liaison, NAEYC Board

Sara Hanhan, University of North Dakota

NAEYC Staff

Heather Biggar, NAEYC Professional Development Specialist, and Marilou Hyson, NAEYC Senior Advisor for Research and Professional Practice

Attachment B
Standard Setting Report

Setting Standards for ELDA K-2

Michael B. Bunch and Kevin B. Joldersma
Measurement Incorporated
FINAL - February 2006

Standard setting for the inventories of ELDA K-2 was accomplished by means of a generalized holistic procedure (cf. Cizek, Bunch, & Koons, 2004) with the assistance of ELL educators in five states. Panelists convened twice by web conference, first for training in the procedure to be used, and later to discuss results of the first round of rating of profiles. Panelists completed all inventory rating off site and sent their ratings via e-mail or fax to Measurement Incorporated for processing.

Procedures

The original plan for standard setting for ELDA K-2 called for a face-to-face two-day meeting of panelists in Durham, North Carolina (see Appendix A). Travel out of state proved to be difficult for many potential panelists; therefore, the procedure was modified to permit web-based training and off-site inventory rating. The web-based training used Netbriefings which supplied both web hosting and conference call hosting for all participants.

Panelists. Panelists were ELL educators and other stakeholders in five participating states: Iowa, Louisiana, Nebraska, Ohio, and South Carolina. A total of 16 individuals participated (9 for the Kindergarten inventories and 7 for the Grades 1-2 inventories). A list of the names of panelists is included in Appendix A.

Performance level descriptors (PLDs). In December 2005, CCSSO prepared updated versions of performance level descriptors for all five levels of each inventory included in the ELDA K-2 assessment. The levels are Pre-functional (1), Beginning (2), Intermediate (3), Advanced (4), and Fully English Proficient (5). These PLDs were a critical part of the training of panelists. Copies of the updated PLDs are included in Appendix A.

Inventories. In a generalized holistic procedure, the task of panelists is to examine student work samples and classify each one as belonging to one of the five performance levels. As the inventories have yet to be administered operationally, and since the operational inventories are considerably different from the ones field tested in 2005, it was necessary to create virtual inventories; i.e., inventories with synthetic scores. MI staff prepared 20-24 such profiles for each inventory, with scores ranging from near 0 to a perfect score. To keep the rating task within reason, some inventories had fewer profiles than there were score points. For example, in Reading (with 42 possible non-zero scores), there were only 24 profiles. Through the middle of the score range (7-35), only the odd score points were represented in the synthetic profiles. Sample inventories prepared for this activity are included in Appendix A.

To create the profiles, psychometricians at MI examined the order of items on each inventory, took into account the relative difficulty of each item (based on ratings by reviewers at the December meeting) and created hypothetical score profiles that matched as closely as possible the performances of actual students. Table 1 shows the total scores for each profile for each inventory. Although there were two sets of inventories (one for Kindergarten and another for grades 1-2), both sets had the same distribution of scores.

Table 1
Scores of Profiles Reviewed: Kindergarten

Profile	Total Score			
	Listening	Reading	Speaking	Writing
1	1	5	1	1
2	2	6	3	3
3	3	7	5	5
4	4	9	7	7
5	5	11	9	9
6	6	13	10	11
7	7	15	11	13
8	8	17	12	15
9	9	19	13	16
10	10	21	14	17
11	11	23	15	18
12	12	25	16	19
13	13	27	17	20
14	14	29	18	21
15	15	31	19	22
16	16	33	20	23
17	17	35	21	24
18	18	36	22	25
19	19	37	23	26
20	20	38	24 (Maximum)	27 (Maximum)
21	21 (Maximum)	39		
22		40		
23		41		
24		42 (Maximum)		

Training. On January 19, Measurement Incorporated hosted a web conference with ELDA staff, staff from participating state departments of education, and ELL educators within the five participating states. Dr. Bunch made a PowerPoint presentation (included in Appendix A) via the web while participants took part in a teleconference. The broadcast lasted just under two hours, during which time panelists' tasks were explained, and panelists took part in a practice exercise. At the end of the training, Dr. Bunch assigned panelists to groups (K or 1-2) and gave the assignment for Round 1. Panelists had until 10 A.M. (Eastern Time) on January 20 to submit completed rating sheets for Round 1.

At 2 P.M. (Eastern Time) on January 20, Dr. Bunch convened a second web-based conference, this time to present the tallied results of Round 1 and to answer questions about the results of that round. During this conference (web-based and teleconference, as on January 19), there was considerable discussion about the meaning of the various PLDs, and their application within the specific contexts of the profiles being rated. At the end of the discussion, Dr. Bunch gave the assignment for Round 2 and gave panelists until Monday, January 23 to return the completed Round 2 profiles.

Logs of the two web conferences are included in Appendix A.

Analyses

Dr. Joldersma compiled the returns from the panelists and created graphs showing where the cut scores would be, using a graphical technique described in the standard-setting plan (See Appendix A). In addition, Dr. Bunch calculated cut scores using a logistic regression technique, common to the Body of Work procedure (Kingston, Kahl, Sweeney, & Bay, 2001) and other holistic procedures. Essentially, the procedure states the relationship between predicted raw score (Y) and the log odds (logistic) function as follows:

$$Y = \ln(P/(1-P)). \tag{1}$$

When $P = .50$, then $1-P$ will also be $.50$, and Y will be the natural log of 1 (i.e., $.50/.50$), which is 0. Thus, to find the point at which the likelihood of being classified in Category C or higher is $.50$, we find the point at which $Y = 0$. But Y can also be modeled in terms of a simple regression equation:

$$Y = a + bX \tag{2}$$

with a representing the intercept, b representing the slope, and X representing raw score.

Results

Results of Round 2 are presented in Tables 2 and 3 for Kindergarten and grades 1-2, respectively. Preliminary cut scores from both graphical and logistic regression analyses are presented along with the maximum possible score for each inventory. Where there are non-

trivial differences between the two sets of results, we offer an explanation. All analyses are included in Appendix B.

Table 2
Round 2 Results for Kindergarten
Cut Scores for Graphical (Graph) and Logistic Regression (LR) Approaches

	Listening		Reading		Speaking		Writing	
Level	Graph	LR	Graph	LR	Graph	LR	Graph	LR
2	4	--	8	8.4	6	8.3	7	7.4
3	9	--	16	19.8	11.5	13.4	16	16.8
4	15	15.1	36	35.7	18.5	18.1	21.5	21.4
5	19	18.8	40	--	22	21.8	26	25.0
Max.	21		42		24		27	

Table 3
Round 2 Results for Grades 1-2
Cut Scores for Graphical (Graph) and Logistic Regression (LR) Approaches

	Listening		Reading		Speaking		Writing	
Level	Graph	LR	Graph	LR	Graph	LR	Graph	LR
2	6	5.7	10	10.3	8	8.3	8	7.4
3	11	10.7	22	21.3	13.5	13.4	16.5	16.8
4	16	16.2	31	31.8	18	18.1	21	21.4
5	19	19.2	39	38.7	22	21.8	25	25.0
Max.	21		42		24		27	

Results were remarkably stable across methods. The blanks in Table 1 refer to instances in which there were no values to enter into the logistic regression. This is actually a good thing. Normally, panelists rate profiles or work samples in such a way that there is a region of disagreement between two regions of agreement. For example, all panelists might agree that profiles with scores below 10 are at Level 2 and that all profiles with scores above 15 are Level 3. Between scores of 10 and 15, there is some disagreement. Logistic regression uses the data from this region of disagreement to calculate cut scores. If all panelists agreed that profiles with scores up to 12 were at Level 2 and all profiles with scores of 13 or above were Level 3, there would be no area of disagreement, and thus no data for the logistic regression. In three instances with the Kindergarten ratings, this is exactly what happened.

The only noticeable differences in results by method were in Reading and Speaking at the Kindergarten level (both at Level 3). In Kindergarten Reading, at the border between Levels 2 and 3, the graphical approach yielded a cut score of 16. However, at points 19 and 21, the number of panelists classifying profiles at Level 2 actually increased. This apparent confusion or

regression on the part of the panelists accounted for the difference in results, inasmuch as the regression attempts to smooth the plots. In Speaking (Kindergarten profiles), the distribution of ratings for Level 2 was quite narrow, while that for Level 3 was quite wide. There were no oddities in the panelists' ratings, however. Again, the logistic regression, in attempting to smooth the curves went further into the Level 3 distribution before settling on a cut score.

Subsequent to the analyses presented in Tables 2 and 3, there was a telephone conference with members of the Technical Advisory Committee to bring resolution to the sets of recommended cut scores. These resolved cuts are shown in Table 4.

**Table 4
Resolved Cut Scores**

Level	Listening		Reading		Speaking		Writing	
	K	1-2	K	1-2	K	1-2	K	1-2
2	4	6	8	10	6	8	7	8
3	9	11	20	22	12	13	16	17
4	15	16	36	31	18	18	21	21
5	19	19	40	39	22	22	26	25
Max.	21	21	42	42	24	24	27	27

Table 4 makes cut score differences between the two sets of inventories stand out; most are in the expected direction (i.e., higher for grades 1-2), a few are slightly lower for grades 1-2, and one is noticeably lower for grades 1-2. We note that the Level 4 cut score for Reading, grades 1-2 is 31 points, compared to 36 points for the same cut score for Kindergarten. We would also point out that while the skills addressed are the same across the two sets of inventories, the actual requirements for score points 1, 2, and 3 are more demanding on the inventory for grades 1-2 than on the Kindergarten inventory. Even with that information in mind, we expect that there will be further discussion of the differences when the spring 2006 administration has been completed and the cut scores are reviewed for a final time.

Discussion

In general, panelists were able to comprehend the task they were to perform, and most completed their rating tasks on time. The synthetic profiles seemed to work well, and panelists were able to transfer the meaning of the PLDs to the profiles they were rating. The cut scores for each level on each inventory are clearly delineated, and there are no instances in which a perfect score is required for Level 5 or in which the lowest cut is dangerously close to 0.

As noted in the previous section, there will be a further review of the cut scores after the spring 2006 administration, per the plan (see Appendix A). At that time, an articulation committee will examine student data, which by that time will have been scaled to match the current scales for

grades 3-12. As was the case in standard setting for grades 3-12, the articulation committee will make recommendations to keep or modify cut scores, based on the overall impact across grades and subjects.

References

- Cizek, G. J., Bunch, M. B., & Koons, H. (2004). Setting performance standards: Contemporary methods. *Educational Measurement: Issues and Practice*, 23 (4), 31-50.
- Kingston, N. M., Kahl, S. R., Sweeney, K., & Bay, L. (2001). Setting performance standards using the body of work method. In G. J. Cizek (Ed.), *Setting Performance Standards: Concepts, Methods, and Perspectives*, Mahwah, NJ: Erlbaum.

Appendix A
Background Materials
(All materials shown in parentheses are available on a separate CD)

- **Standard Setting Plan** ([AppA\StdSet.doc](#))
- **Participant List** ([AppA\Panelists.xls](#))
- **Performance Level Descriptors** ([AppA\PLDs.doc](#))
- **Web-Based Training Presentation** ([AppA\K-2.ppt](#))
- **Log of January 19 Conference Call** ([AppA\eConference.doc](#))
- **Log of January 20 Conference Call** ([AppA\Jan20Conf.doc](#))
- **Sample Profile 1** ([AppA\L12001.doc](#))
- **Sample Profile 2** ([AppA\L12006.doc](#))
- **Sample Profile 3** ([AppA\L12018.doc](#))
- **Instructions for Completing Round 1** ([AppA\Instructions for Completing.doc](#))
- **Instructions for Completing Round 2** ([AppA\Instructions for Round 2.doc](#))
- **Kindergarten Profiles** ([AppA\ProKR2.doc](#))
- **Grades 1-2 Profiles** ([AppA\Pro12R2.doc](#))
- **Evaluation Form** ([AppA\Eval.doc](#))

Appendix B
Data Analyses
(Both spreadsheets have been provided to CCSSO on CD.)

- Round 1 ([AppB\Appendix B1.xls](#))
- Round 2 ([AppB\Appendix B2.xls](#))

ELDA K-2 Performance Level Descriptors

LISTENING

Entry into 5 (FEP)

Students at this level understand most grade-level appropriate content-area and school/social speech. They understand the main ideas and relevant details of extended discussions or oral presentations on a range of familiar and unfamiliar topics comparable to a native English speaker at the same grade level. They are capable of making interpretations of what they hear. They understand most of the complex structures of spoken English relative to their grade level. They have a broad range of vocabulary, including idiomatic language, relating to both content areas and school/social environments.

Entry into 4 (Advanced)

Students at this level understand conversations in most school/social settings. They understand main ideas and significant relevant details of extended discussions or presentations on familiar and relevant academic topics. They are able to comprehend conversations and orally-delivered texts involving description and narration in different time frames or conditions. They understand most of the basic language forms of spoken English including timeless conditionals and sentences using clauses and phrases. They are able to understand cohesive devices to follow the sequence in an oral presentation or text. They comprehend most grade-level vocabulary and idioms, especially school/social environments, and are beginning to develop a wide range of academic vocabulary related to content areas, with limited supports such as visuals and rephrasing. They understand multiple meanings of words and can use context clues to understand messages.

Entry into 3 (Intermediate)

Students at this level understand sentence-length statements and questions that include recombinations of learned language structures and on a variety of social and academic topics. They understand simple and compound sentences. They understand time through the use of simple tenses that may not be supported by adverbials of time. They are able to understand multi-step directions. They also understand the difference between statements and questions by intonation, word order, and interrogative words. They understand and are able to identify main ideas and some details from conversations and simple/age appropriate orally-delivered text, usually with visual supports in familiar communicative situations and in academic content areas. They begin to interpret meaning from conversations and orally-delivered text, making predictions and drawing conclusions. They understand some idioms, mostly related to school/social environments, and have key vocabulary from content areas. They are aware of cohesive devices but may not be able to use them to follow the sequence of thought in an oral text.

Entry into 2 (Beginning)

Students at this level understand simple, short statements and questions on a well-known topic within a familiar context. Tense is understood through the use of adverbials or situation rather than inflectional endings. They are able to follow simple multi-step directions. They identify the main idea and some details of short conversations or simple orally-delivered text on a familiar

topic. They understand basic grammatical structures and vocabulary in the school and social environment. Students at this level still need frequent repetition and rephrasing. They understand what they have heard but not variations or recombinations of what they have heard.

0→1(Pre-functional)

Students at this level may understand some isolated words (particularly school and social environment vocabulary), some high frequency social conventions, and simple (single word or short phrase) directions, commands and questions. They rely on non-verbal cues such as gestures and facial expressions and require frequent repetition and rephrasing to understand spoken language. They need strong situational support to understand most oral language.

SPEAKING

Entry into 5 (FEP)

Students who are ready to enter Level 5, Fully English Proficient, can supply coherent, unified and appropriately sequenced responses to an interlocutor. They use a variety of devices to connect ideas logically. They understand and can use a range of complex and simple grammatical structures, as appropriate for topic and type of discourse. Their grammar and vocabulary is comparable to that of a minimally proficient native English speaker—grammar errors very seldom impede communication and their range of school-social and academic vocabulary allows a precision of speech comparable to a native English speaker. They can effectively engage in non-interactive speech. They can use language effectively to connect, tell, expand, and reason. They show flexibility, creativity and spontaneity in speech in a variety of contexts.

Entry into 4 (Advanced)

Students entering proficiency Level 4, the Advanced level, are able to restructure the language they know to meet the creative demands of most social and academic situations. They can supply mostly coherent, unified and appropriately sequenced responses to an interlocutor. They use some devices to connect ideas logically and they use a range of grammatical structures. They make some errors in modality, tense, agreement, pronoun use, and inflections. Students have sufficient vocabulary to communicate in non-academic situations and most academic ones. They can engage in extended discussions. They can often use language to connect, tell and expand on a topic; and can begin to use it to reason. They are fluent but may still hesitate in spontaneous in communicative situations.

Entry into 3 (Intermediate)

Students entering proficiency level 3, the Intermediate level, are no longer wholly dependent on practiced, memorized, or formulaic language. They restructure learned language to communicate on a range of subjects. Their speech is still marked by errors in modality, tense, agreement, pronoun use, and inflections. These errors seldom interfere with communication in simple sentences, but do interfere in complex constructions. Intermediate level students are limited in vocabulary, especially academic vocabulary. They can retell, describe, narrate, question, and give instructions, although they lack fluidity and fluency when not using practiced or formulaic language. They often use language to connect, tell and sometimes to expand on a known topic.

Entry into 2 (Beginning)

Students who are just entering proficiency level 2, the beginning level, predominantly use formulaic patterns and memorized phrases. When they deviate from formulaic language, their speech imitates telegraphic language due to the omission of some meaningful linguistic components. Their language is also marked by the lack of tense, number, and agreement. They may use some very simple transitional markers, usually “and” to link ideas. They rely on schemata in L1. Their school-social vocabulary is limited to key words and they have little or no academic vocabulary. They respond to questions usually with one or two-word answers. They can connect and tell on a known topic.

0→1 (Pre-functional)

Students at this level may say or repeat common phrases, words and formulaic language. They may be able to provide some basic information in response to requests and questions. They can ask one or two-word questions without regard to structure and intonation.

READING

Entry into 5 (FEP)

Students at this level participate in reading activities with little teacher support at a level comparable to their English-speaking peers. They read for different purposes across a variety of text types. They have an increasing range of receptive nonacademic and academic vocabulary that allows them to read with greater fluency. They understand multiple word meanings. They have greater comprehension as a result of their increasing control of the structures of English. They can make connections between what they read and other experiences and tasks.

Entry into 4 (Advanced)

Students at this level can read familiar text with little teacher or visual support. However, they still need those supports when reading to comprehend unfamiliar text. They can apply their phonemic awareness skills to read more complicated text. They have oral fluency and use self-monitoring and self-correction strategies when necessary. They use pre, during and post reading strategies but still need teacher prompting to use these skills. They can identify all story elements and can recognize cause and effect relationships in the texts they read. They make connections between the texts they read and themselves, the world, and other texts. They comprehend text in read aloud and can participate in the majority of read aloud activities. They are beginning to read across text types and apply what they read to other activities.

Entry into 3 (Intermediate)

Students at this level are developing phonemic awareness skills that allow them to read single words and simple text with comprehension. Reading is aided by visual and teacher supports. At this stage oral reading is hesitant and difficult to understand due to a lack of oral language proficiency. These students have a small repertoire of high frequency words. They are beginning to use simple reading strategies and to make self, world, and other text connections to the text they are reading. They comprehend simple sentence structure and sentences with simple compounding. They recognize that words serve different functions, have multiple meanings, and

have both synonyms and antonyms. In read aloud, with teacher support, they can identify some story elements and retell the majority of the story.

Entry into 2 (Beginning)

Students at this level begin to identify the names of both upper and lower case letters of the alphabet. They use juncture to identify where words begin and end. They begin to recognize that words serve different functions (e.g. nouns, verbs). They can follow multi-step directions depicted graphically. During read aloud they get meaning primarily from pictures and the teacher's tone of voice and gestures.

0→1 (Pre-functional)

Students at this level demonstrate an understanding of concepts of print (e.g., front-to-back, top-to-bottom, left-to-right) and begin to track print. They can distinguish letters from other symbolic representations. They can follow one-step directions depicted graphically. They can imitate the act of reading (e.g. holding a book and turning pages); however, they get meaning only through pictures.

WRITING

Entry into 5 (FEP)

Students at this level participate in writing activities with no teacher support. They write across all text types. They edit for sentence-level structure, spelling, and mechanics and revise for content, organization and vocabulary. They can use complex sentence structures, with some errors, and can edit for syntax and grammar. They have a range of nonacademic and academic vocabulary that allows for precision, and they begin to use nuanced and alternative word meanings. They employ subtleties for different audiences and purposes. They can use appropriate writing conventions with some errors that do not affect comprehensibility.

Entry into 4 (Advanced)

Students at this level participate in writing activities with minimal teacher support. They are able to restructure in writing the language they know to meet the creative demands of most social and academic situations. They can write mostly coherent, unified, and appropriately sequenced sentences. They use devices to connect ideas logically. They use a range of grammatical structures and can switch appropriately from one tense to another as required by the time frame of their text. They make some errors in modality, tense, agreement, pronoun use, and inflections. Students have a strong BICS vocabulary and a functional academic vocabulary that allows them to participate meaningfully in content classes. They write using all text types, at a developmentally appropriate level. They edit for sentence-level structure, spelling, and mechanics and revise for content, organization and vocabulary.

Entry into 3 (Intermediate)

Students at this level participate in writing activities with some teacher support. They can write simple and compound sentences and are beginning to write with phrases. They use simple tenses, number, and agreement with random errors. They use transition words to link sentences and order these in a developmentally appropriate manner. They begin to edit for sentence-level structure, spelling and mechanics and revise for content, organization and vocabulary, usually with the support of the teacher. They have a good range of BICS vocabulary and are beginning to use more academic content-specific words. They write mostly descriptive, expository, procedural, and narrative text. Their writing is less dependent on visual supports, shared experiences, and scaffolding.

Entry into 2 (Beginning)

Students at this level participate in writing activities by drawing pictures or dictating words. They are able to write connected words and short telegraphic sentences. They are able to revise or edit their writing with teacher support. Their writing is marked by the lack of tense, number, and agreement. They may use some simple transitional markers, usually “and” to link ideas. Their vocabulary reflects what they can say orally. They make frequent errors in mechanics such as punctuation and capitalization. They write mostly descriptive, expository, and procedural text. Their writing is most effective when supported by a visual, a shared experience, or scaffolding.

0→1 (Pre-functional)

Students at this level participate in writing activities by drawing pictures. They may be able to copy letters or form them from memory and may be able to copy some words. They can imitate the act of writing (e.g. scribbling); however, their text does not transmit a message. They may attempt to apply some writing conventions but do so inappropriately or do so correctly only when copying.

Attachment C
Evaluation Comments

Question #	Selection	Comments regarding question
	1 D	Came very late.
	2 SD	Just think of the number of students & the work of the teacher!
	2 A	We had to add our own name labels.
	3 SD	It should say, "Not for Kindergarten."
	4 SD	Kindergarten section not well defined -- doesn't need 1st and 2nd grade section.
	4 D	Speaking photos are very unclear.
	4 D	Not for students
	5 A	Tasks are one thing. The way to find out if the students fulfill the tasks is another thing!
	5 D	Some only appropriate for older students.
	5 D	Especially with rows 25-28
	6 SD	Where is the direction? The notes at the end? They are not at all helpful.
	6 D	Everything has to be "studied."
	7 A	The tasks are consistent with the core curriculum. What's wrong?
	8 A & D	Very time consuming + tedious
	8 D	Use of "phrase" / The sample was not in test booklet
	9 A & D	9, 15 + 16 were N/A -- shouldn't be in there.
	9 D	"Bean Pots" should be sentences / Many students questioned "the new student in class" -- they didn't have one.
	10 A & D	tedious + time consuming
	10 D	"Phrase" example not in book
	11 A	#6 should have a step between 1 + 2
	11 D	"Lunchtime rule" confused students
	12 D	Some students didn't know what to do. They needed to be simplified. Directions were too wordy for the students.
	12 A & D	All very tedious + time consuming
	12 D	No need to repeat twice

13	SD	Even parents may not remember when their child first arrived in the U.S.
13	D	Type of ESL program was a ? Also, some of the info like what schooling in the "other country" we do not know.
13	A	There's no explanation for how long someone has been in the U.S. -- does being born here count?
13	D	Questions 12, 13, 16, 17 not always available
13	A	Too much info
14	D	The book was very unclear for Kindergarten. If I did not have skills test results, IPT results, and if I did not have my students 3 periods a day, I could <u>not</u> administer this test in the allotted time!!! It will take a new teacher with 40 some tests to administer <u>FOREVER!!!</u> This test is is <u>ludicrous</u> for districts with large groups of ELL's.
14	A	At this time of yr., even pulling together the info. for the inventories is too overwhelming for the tchrs. I work with.
14	D	Barely -- test window too near end of school
15	SD	Only because I only had 8 students; not 40 as usual.
15	D	This will be difficult to observe with 20-30 students
15	SD	Only administered Reading Inventory to 1 student.
15	D	Documentation difficult to compile
16	D	With working with my students I already know how well they do, but if I was an outside teacher/or other teacher who doesn't know the students it would be hard to observe 25 students in 2 weeks.
16	SD	I was able to make meaningful entries with IPT + Skills Test results <u>and</u> daily/yearly observations.
16	A/D	A = classroom teacher <u>only</u> ; D = if done by outsider
16	D	This will be difficult to observe with 20-30 students
16	none	I did the best I could with the time I had. Longer observation times would provide more reliability.
16	A	Documentation difficult to compile
16	SA	I only had 1 student. If a teacher had several I think she needs a day off to do it. This takes about 3 hours.

