State Implementation of NCLB Policies and Interpretation of the NAEP Performance of English Language Learners

Richard P. Durán
Gevirtz Graduate School of Education
University of California, Santa Barbara

Commissioned by the NAEP Validity Studies (NVS) Panel
December 2006

George W. Bohrnstedt, Panel Chair
Frances B. Stancavage, Project Director

The NAEP Validity Studies Panel was formed by the American Institutes for Research under contract with the National Center for Education Statistics. Points of view or opinions expressed in this paper do not necessarily represent the official positions of the U.S. Department of Education or the American Institutes for Research.

The NAEP Validity Studies (NVS) Panel was formed in 1995 to provide a technical review of NAEP plans and products and to identify technical concerns and promising techniques worthy of further study and research. The members of the panel have been charged with writing focused studies and issue papers on the most salient of the identified issues.

Panel Members:

Project Director:

Frances B. Stancavage
American Institutes for Research

Project Officer:

Janis Brown
National Center for Education Statistics

For Information:

NAEP Validity Studies (NVS)
American Institutes for Research
1070 Arastradero Road, Suite 200
Palo Alto, CA 94304-1334
Phone: 650/ 843-8192
Fax: 650/ 858-0958

CONTENTS

LIST OF EXHIBITS

Introduction

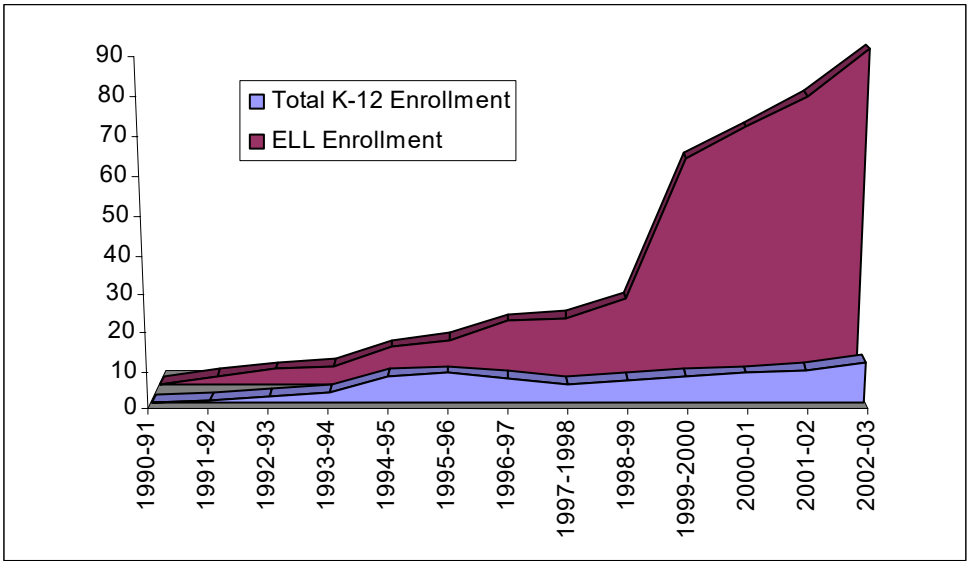
Since its inception more than 40 years ago, the National Assessment of Educational Progress (NAEP) has served as the key indicator for the educational achievement of the nation's youth (Jones and Olkin, 2004). During the same period, the initial passage and successive reauthorization of the Elementary and Secondary Education Act (ESEA) of 1965 has grappled with improving educational outcomes for all students—in particular, those from low-income backgrounds and backgrounds associated with at-risk factors affecting educational outcomes. In this context, NAEP has come to be an even more important and essential resource for uncovering evidence of progress toward national goals associated with the ESEA.

The 2001 No Child Left Behind (NCLB) reauthorization of the ESEA requires that states receiving Title I, Part A funds participate in both NAEP state reading and mathematics assessments biennially at grades 4 and 8 beginning in 2002-03. While not required formally, other states and education agencies are encouraged to interpret the performance of students on state mandated assessments in light of their performance on NAEP. NCLB also requires that states, school districts, and schools report the academic achievement and Adequate Yearly Progress (AYP) of mandated subgroups of students who historically have shown patterns of low educational achievement and attainment in the aggregate. Limited English proficient students (LEP), equivalently termed "English language learners" (ELLs) in this report are one of these targeted subgroups.[1] In turn, state NAEP now reports reading and mathematics scores for ELLs.

The number of students classified as ELL in the 50 states and the District of Columbia has grown significantly over the past decade. Based on U.S. Department of Education state surveys and enrollment totals from the National Center for Education Statistics (NCES), there are 5.5 million students, who speak more than 400 different languages, in the ELL student population enrolled in pre-K through grade 12. Eighty percent of these students speak Spanish as a first language (U.S. Department of Education, retrieved 2-26-06). In 1997-98, the number of ELL students in the same grade range was estimated at somewhat less than 3.5 million children; this represents a growth rate of more than 55 percent during this period. Between 1989-90 and 2003-04, ELL enrollment in K-12 has more than doubled (from 2,030,451 to 5,013,539 students). As shown in Exhibit 1, while the total enrollment of all K-12 students grew about 20 percent between 1989-90 and 2003-04, the ELL student population in grades K-12 grew by about 147 percent. At present, ELL students constitute approximately 10 percent of the 49.5 million children enrolled in grades K-12 nationwide (National Clearinghouse for English Language Acquisition and Language Instruction Educational Programs, retrieved 2-26-06).

[1] In the NCLB legislation, the term "limited English proficient" (LEP) is used to refer to students from a non-English first language background who are in the process of acquiring initial proficiency in the English language. Alternatively, the term "English language learner" (ELL) is in common usage.

Exhibit 1. Relative Growth in ELL and Total Enrollment in U.S. Schools, 1989-90 to 2003-04



Source: National Clearinghouse for English Language Acquisition and Language Instruction Educational Programs. (retrieved 2-26-06 from http://www.ncela.gwu.edu/expert/faq/08leps.html).

This report outlines a number of critical issues that should be addressed in order to allow states to explore and understand relationships between the performance of ELL students on NAEP and on state assessments in this policy context. The results of this study can be useful to a variety of education stakeholders who are interested in improving the utility of NAEP for examining the performance of ELL students.

The unique characteristics of ELL students raise many interesting and important challenges from both a policy and statistical measurement perspective. For one thing, the policy context and validity issues are complicated because states' enactment of NCLB provisions is itself a developmental process. In addition, as NCLB is implemented, adjustments have been made in its provisions to assist states in complying with its intent.

The measurement context is also complex. NAEP and state assessments in reading and mathematics are not developed for exactly the same purposes, and they do not have exactly the same measurement properties. As will be discussed in this report, attention to a formal statistical study of the relationships between ELL scores on NAEP and state assessments will help to inform the validity rationale underlying attempts to compare results across these assessments.

In order to enhance the discussion and make concrete the issues under investigation, this report briefly reviews both provisions for ELL participation in NAEP and NCLB provisions for ELL assessment. In addition, the results of an exploration of NCLB policies and practices in four states (California, New York, Texas and Washington) are examined. These four states served an estimated 51 percent of all pre-K through grade 12 ELL students nationwide in 2001-02. Exploration of issues for the four target states helps illustrate key validity challenges faced by states as they consider investigation of relationships between NAEP scores and state assessment scores for ELLs under NCLB *for their individual state*.

Finally, consideration is given to next steps that the NAEP program might take to improve states' use of NAEP scores as part of their analysis of progress in attaining NCLB goals. The discussion also includes consideration of challenges in examining ELL student progress toward attaining NCLB goals across states. The latter theme is explored in this

report in terms of strategies that the NAEP program might begin to pursue to assist states in better standardizing NAEP participation procedures and practices for ELL students. Also discussed are ways in which the NAEP program might assist states in developing common strategies for studying relationships between NAEP and state assessment performance for ELL students. The latter discussion includes attention to variation in states' readiness to explore relationships between ELL students' performance on NAEP and state assessments based on state population characteristics, NAEP participation, and state accountability system design.

NAEP Provisions for ELL Participation

Under current NAEP policy, all ELL students who have received academic instruction in English for 3 years or more (including the current year) are to be included in NAEP assessments if selected to participate (NCES, 2006). ELL students who have received instruction in English for fewer than 3 years also should be included, if selected, unless school personnel judge them to be incapable of participating in the assessment in English. In order to enhance participation of ELL students in NAEP, states are permitted to provide certain assessment accommodations (see Exhibit 2). ELL students (and students with disabilities) may use the same accommodation in NAEP assessments that they use in their usual classroom assessments. However, accommodations that make it impossible to measure a target skill or ability are disallowed, as are accommodations not within the capacity of the NAEP program. As an example of the former, and particularly relevant to ELL students, states are not permitted to have NAEP administrators read aloud passages or questions to students on the NAEP reading achievement test.

ELL students in grades 4 and 8 in 1996, 2000, 2003, and 2005 had the option of taking the NAEP mathematics assessment in a bilingual book format that presented problems in a side-by-side, Spanish-English format.

Exhibit 2. Accommodations Frequently Provided for English Language Learners (ELL), Showing Which Accommodations are Permitted by NAEP

Frequently Provided Accommodations for English Language Learners	Permitted by NAEP?
Native language version of test	No
Bilingual version of test (Spanish/English)	No (except for mathematics and science)
Bilingual word lists or glossaries (Spanish/English)	No (except for science)
Bilingual dictionary without definitions	Yes[1] (except for reading)
Directions translated aloud into native language or presented by audiotape	No
Student's oral or written responses translated into written English	No
Passages, other stimulus materials, or test questions read aloud in English or presented by audiotape	Yes (except for reading)
Directions read aloud in English or presented by audiotape*	Yes
Passages, other stimulus material, or test questions translated aloud into native language or presented by audiotape **	No
Small group	Yes
One-on-one (tested individually)	Yes
Extended time	Yes
Preferential seating	Yes

[1] Not provided by NAEP, but school, district, or state may provide after fulfilling NAEP security requirements.
* Standard NAEP practice. Not considered an accommodation.
** For Spanish/English bilingual mathematics and science, this would be standard NAEP practice. Not allowed otherwise.
SOURCE: National Center for Education Statistics (retrieved 10-23-06 from http://nces.ed.gov/nationsreportcard/about/inclusion.asp).

Historically, the intent of successive NAEP inclusion policies and increasing use of accommodations has been to maximize the representation of NAEP results for the nation's youth at the assessment's target grade levels. While this goal includes reporting accurate and valid national and state estimates of NAEP performance for a variety of ethnic-racial groups and by gender, the implementation of NCLB and its requirement that states track the achievement progress of ELL students has raised additional challenges for NAEP, given its high standards for quality of data reporting, concern for states' interpretations of their NAEP scores, and concern for ensuring consistency of data reporting and interpretation across states. Before discussing these challenges, this report will focus on what NAEP tells us about the reading and mathematics achievement of ELL students who are able to be assessed by NAEP.

2005 NAEP Data on Reading and Mathematics Achievement of ELL Students

Exhibits 3 through 6 present a summary of performance results on the NAEP 2005 reading and mathematics assessments disaggregated by ELL, non-ELL, and former ELL status for the nation as a whole and by jurisdiction.[2] Caution needs to be exercised in interpreting these results because the only ELL students included were those who were not excluded from the assessment due to inability to take an English version of the assessment or inability to take the NAEP assessment with a permitted accommodation. As will be

[2] Former ELL students include students who passed their states' English language proficiency examinations in the past 2 years.

discussed throughout the report, caution also needs to be exercised about comparing the results for ELL students because operationalization of the ELL classification itself and its interaction with participation in NAEP varies from state-to-state, which make national estimates and cross-state comparisons involving ELL students subject to systematic sources of error that go uncorrected (Haertel, 2003).

As the data indicate, the national average scale score for ELL students participating in the 2005 NAEP reading assessment was lower than the average scale score for non-ELL students by 33 points at grade 4 and 39 points at grade 8. In mathematics, the national average scale score for ELL students was lower than the average scale score for non-ELL students by 23 points at grade 4 and 36 points at grade 8. Former ELL students performed much like non-ELL students in both subjects and at both grade levels.

The effect size of these differences can only be estimated by adjusting for a potential bias in the means for ELL students. This requires making additional assumptions about the achievement characteristics of excluded ELL students relative to included ELL students and students as a whole. Nonetheless, the data show that ELL students perform at noticeably lower levels of performance, as expected.

Inspection of differences in the percentage of ELL and non-ELL students classified as "Below Basic," "At or Above Basic," or "At or Above Proficient" in reading and mathematics at grade 4 and 8 shows a consistent pattern. Across both subjects and both grade levels, the percentages of ELL students participating in the 2005 NAEP assessments who scored at or above the proficient level ranged from 4 percent for grade 8 reading to 11 percent for grade 4 mathematics. By comparison 30 to 38 percent of non-ELL students scored in this range. Furthermore, for reading at both grades and mathematics at grade 8, the modal achievement classification for ELL students was "Below Basic," while the corresponding modal classification for non-ELL students was "At or Above Basic." (Both groups had modal scores "At or Above Basic" for grade 4 mathematics.)

Attention will now turn to NCLB provisions regarding inclusion and assessment of ELL students in state assessment systems.

Exhibit 3. Average Reading Scales Scores and Achievement-level Results, by English Language Learners (ELL), Grade 4 Public Schools: By State, 2005

	ELL					Non-ELL					Formerly ELL				
			Percentage of students					Percentage of students					Percentage of students		
State/Jurisdiction	Percentage of all students	Average scale score	Below *Basic*	At or above *Basic*	At or above *Proficient*	Percentage of all students	Average scale score	Below *Basic*	At or above *Basic*	At or above *Proficient*	Percentage of all students	Average scale score	Below *Basic*	At or above *Basic*	At or above *Proficient*
Nation (public)	9	187	73	27	7	90	220	34	66	32	1	217	38	62	26
Alabama	1	‡	‡	‡	‡	99	208	47	53	23	#	‡	‡	‡	‡
Alaska	18	177	77	23	7	82	219	34	66	31	#	‡	‡	‡	‡
Arizona	18	175	81	19	4	82	214	41	59	28	1	‡	‡	‡	‡
Arkansas	3	205	53	47	17	97	217	37	63	30	#	‡	‡	‡	‡
California	31	183	77	23	5	66	217	38	62	29	2	221	33	67	30
Colorado	10	191	71	29	7	90	227	26	74	40	1	‡	‡	‡	‡
Connecticut	4	193	66	34	8	96	227	28	72	40	#	‡	‡	‡	‡
Delaware	3	206	53	47	16	97	226	26	74	35	#	‡	‡	‡	‡
Florida	6	193	68	32	7	91	222	33	67	32	4	209	50	50	20
Georgia	2	182	80	20	4	98	215	41	59	27	#	‡	‡	‡	‡
Hawaii	8	183	78	22	6	92	212	44	56	25	#	‡	‡	‡	‡
Idaho	8	191	69	31	6	92	225	28	72	35	1	‡	‡	‡	‡
Illinois	7	176	82	18	4	92	220	34	66	32	#	‡	‡	‡	‡
Indiana	2	‡	‡	‡	‡	98	218	36	64	31	#	‡	‡	‡	‡
Iowa	3	‡	‡	‡	‡	97	222	31	69	34	#	‡	‡	‡	‡
Kansas	6	195	65	35	9	94	222	32	68	34	#	‡	‡	‡	‡
Kentucky	1	‡	‡	‡	‡	99	220	35	65	31	#	‡	‡	‡	‡
Louisiana	1	‡	‡	‡	‡	99	209	47	53	20	#	‡	‡	‡	‡
Maine	1	‡	‡	‡	‡	99	225	29	71	35	#	‡	‡	‡	‡
Maryland	2	‡	‡	‡	‡	98	221	35	65	33	#	‡	‡	‡	‡
Massachusetts	5	198	61	39	11	94	233	20	80	46	2	208	53	47	16
Michigan	2	‡	‡	‡	‡	98	219	36	64	32	#	‡	‡	‡	‡
Minnesota	6	199	57	43	10	94	227	27	73	40	#	‡	‡	‡	‡
Mississippi	1	‡	‡	‡	‡	99	205	52	48	18	#	‡	‡	‡	‡
Missouri	1	‡	‡	‡	‡	99	222	32	68	33	#	‡	‡	‡	‡
Montana	3	‡	‡	‡	‡	97	226	27	73	37	#	‡	‡	‡	‡
Nebraska	6	187	74	26	4	93	224	30	70	36	1	‡	‡	‡	‡
Nevada	14	176	83	17	3	86	212	42	58	23	#	‡	‡	‡	‡
New Hampshire	2	‡	‡	‡	‡	98	228	25	75	39	#	‡	‡	‡	‡
New Jersey	2	‡	‡	‡	‡	98	224	31	69	38	#	‡	‡	‡	‡
New Mexico	19	182	76	24	5	81	213	42	58	24	#	‡	‡	‡	‡
New York	5	186	75	25	3	90	225	29	71	35	6	222	33	67	29
North Carolina	6	192	70	30	7	93	219	36	64	31	1	215	40	60	28
North Dakota	1	‡	‡	‡	‡	99	225	28	72	36	#	‡	‡	‡	‡
Ohio	1	‡	‡	‡	‡	99	223	31	69	35	#	‡	‡	‡	‡
Oklahoma	4	192	66	34	8	95	215	38	62	26	#	‡	‡	‡	‡
Oregon	12	187	73	27	7	88	221	33	67	33	#	‡	‡	‡	‡
Pennsylvania	2	196	58	42	16	98	223	31	69	36	#	‡	‡	‡	‡
Rhode Island	6	172	85	15	2	94	219	35	65	31	#	‡	‡	‡	‡
South Carolina	1	‡	‡	‡	‡	99	213	42	58	26	#	‡	‡	‡	‡
South Dakota	3	178	85	15	2	97	224	29	71	34	#	‡	‡	‡	‡
Tennessee	2	‡	‡	‡	‡	98	215	40	60	27	#	‡	‡	‡	‡
Texas	10	196	65	35	8	87	222	32	68	32	2	‡	‡	‡	‡
Utah	9	191	67	33	11	90	225	29	71	37	1	‡	‡	‡	‡
Vermont	1	‡	‡	‡	‡	98	227	28	72	39	#	‡	‡	‡	‡
Virginia	6	214	40	60	22	94	227	27	73	38	#	‡	‡	‡	‡
Washington	8	191	70	30	6	92	226	27	73	38	#	‡	‡	‡	‡
West Virginia	1	‡	‡	‡	‡	99	215	39	61	26	#	‡	‡	‡	‡
Wisconsin	5	202	58	42	14	95	222	31	69	34	#	‡	‡	‡	‡
Wyoming	4	190	71	29	4	95	225	27	73	36	#	‡	‡	‡	‡
Other jurisdictions															
District of Columbia	5	177	80	20	4	95	191	66	34	12	#	‡	‡	‡	‡
DoDEA[1]	6	203	56	44	11	94	228	24	76	37	#	‡	‡	‡	‡

#The estimate rounds to zero.
‡Reporting standards not met. Sample size in insufficient to permit a reliable estimate.
[1] Department of Defense Domestic Dependent Elementary and Secondary Schools
NOTE: ELL = English language learners. Formerly ELL = students who passed their state's English-language proficiency examination within the past 2 years. The results for English language learners are based on students who were assessed and cannot be generalized to the total population of such students. Details may not sum to totals because of rounding.
SOURCE: Perie, Grigg, and Donahue (2005).

Exhibit 4. Average Reading Scales Scores and Achievement-level Results, by English Language Learners (ELL), Grade 8 Public Schools: By State, 2005

	ELL					Non-ELL					Formerly ELL				
			Percentage of students					Percentage of students					Percentage of students		
State/Jurisdiction	Percentage of all students	Average scale score	Below *Basic*	At or above *Basic*	At or above *Proficient*	Percentage of all students	Average scale score	Below *Basic*	At or above *Basic*	At or above *Proficient*	Percentage of all students	Average scale score	Below *Basic*	At or above *Basic*	At or above *Proficient*
Nation (public)	5	224	71	29	4	93	263	27	73	30	2	255	34	66	20
Alabama	1	‡	‡	‡	‡	99	252	37	63	22	#	‡	‡	‡	‡
Alaska	14	234	59	41	8	86	263	25	75	29	#	‡	‡	‡	‡
Arizona	12	225	75	25	3	87	259	30	70	26	1	‡	‡	‡	‡
Arkansas	1	‡	‡	‡	‡	99	258	31	69	26	#	‡	‡	‡	‡
California	20	222	74	26	3	75	258	32	68	25	5	258	30	70	20
Colorado	5	229	69	31	5	94	267	22	78	33	1	‡	‡	‡	‡
Connecticut	2	‡	‡	‡	‡	98	265	25	75	34	#	‡	‡	‡	‡
Delaware	2	‡	‡	‡	‡	98	267	19	81	31	#	‡	‡	‡	‡
Florida	4	221	73	27	5	95	257	32	68	26	2	250	43	57	19
Georgia	2	‡	‡	‡	‡	98	257	33	67	25	#	‡	‡	‡	‡
Hawaii	5	212	85	15	1	95	250	40	60	19	#	‡	‡	‡	‡
Idaho	4	241	48	52	12	95	265	23	77	33	#	‡	‡	‡	‡
Illinois	2	227	66	34	6	98	264	25	75	31	#	‡	‡	‡	‡
Indiana	1	‡	‡	‡	‡	99	261	27	73	29	#	‡	‡	‡	‡
Iowa	1	‡	‡	‡	‡	99	268	20	80	34	#	‡	‡	‡	‡
Kansas	2	‡	‡	‡	‡	98	267	22	78	35	#	‡	‡	‡	‡
Kentucky	1	‡	‡	‡	‡	99	264	25	75	31	#	‡	‡	‡	‡
Louisiana	1	‡	‡	‡	‡	99	253	36	64	20	#	‡	‡	‡	‡
Maine	1	‡	‡	‡	‡	99	270	18	82	38	#	‡	‡	‡	‡
Maryland	#	‡	‡	‡	‡	100	261	30	70	30	#	‡	‡	‡	‡
Massachusetts	2	222	74	26	2	97	275	16	84	45	1	238	56	44	6
Michigan	2	‡	‡	‡	‡	98	261	27	73	29	#	‡	‡	‡	‡
Minnesota	5	‡	‡	‡	‡	94	271	18	82	39	#	‡	‡	‡	‡
Mississippi	#	‡	‡	‡	‡	100	251	40	60	19	#	‡	‡	‡	‡
Missouri	#	‡	‡	‡	‡	100	265	24	76	31	#	‡	‡	‡	‡
Montana	4	230	67	33	2	96	271	16	84	38	#	‡	‡	‡	‡
Nebraska	2	‡	‡	‡	‡	97	268	19	81	35	1	‡	‡	‡	‡
Nevada	10	221	76	24	2	89	257	32	68	25	1	‡	‡	‡	‡
New Hampshire	1	‡	‡	‡	‡	99	270	19	81	38	#	‡	‡	‡	‡
New Jersey	1	‡	‡	‡	‡	98	270	19	81	38	1	‡	‡	‡	‡
New Mexico	13	224	70	30	3	87	255	34	66	22	#	‡	‡	‡	‡
New York	3	221	74	26	4	88	267	22	78	36	9	257	32	68	22
North Carolina	3	236	57	43	7	97	259	30	70	28	1	‡	‡	‡	‡
North Dakota	1	‡	‡	‡	‡	99	270	16	84	36	#	‡	‡	‡	‡
Ohio	#	‡	‡	‡	‡	100	267	22	78	36	#	‡	‡	‡	‡
Oklahoma	3	‡	‡	‡	‡	97	260	27	73	26	#	‡	‡	‡	‡
Oregon	7	235	58	42	9	93	265	24	76	34	#	‡	‡	‡	‡
Pennsylvania	1	‡	‡	‡	‡	99	267	22	78	36	#	‡	‡	‡	‡
Rhode Island	3	215	74	26	3	97	263	27	73	30	#	‡	‡	‡	‡
South Carolina	1	‡	‡	‡	‡	99	257	33	67	25	#	‡	‡	‡	‡
South Dakota	2	‡	‡	‡	‡	98	269	17	83	36	#	‡	‡	‡	‡
Tennessee	1	‡	‡	‡	‡	99	259	29	71	26	#	‡	‡	‡	‡
Texas	6	216	79	21	2	93	261	28	72	28	1	243	47	53	9
Utah	6	234	60	40	7	94	264	25	75	31	1	‡	‡	‡	‡
Vermont	1	‡	‡	‡	‡	99	269	21	79	38	#	‡	‡	‡	‡
Virginia	2	‡	‡	‡	‡	98	268	21	79	36	#	‡	‡	‡	‡
Washington	4	224	70	30	5	96	267	22	78	36	#	‡	‡	‡	‡
West Virginia	1	‡	‡	‡	‡	99	255	33	67	22	#	‡	‡	‡	‡
Wisconsin	2	‡	‡	‡	‡	98	267	23	77	35	#	‡	‡	‡	‡
Wyoming	3	242	50	50	8	97	269	18	82	37	#	‡	‡	‡	‡
Other jurisdictions															
District of Columbia	2	‡	‡	‡	‡	98	238	55	45	12	#	‡	‡	‡	‡
DoDEA[1]	3	‡	‡	‡	‡	97	271	16	84	38	#	‡	‡	‡	‡

#The estimate rounds to zero.
‡Reporting standards not met. Sample size in insufficient to permit a reliable estimate.
[1] Department of Defense Domestic Dependent Elementary and Secondary Schools
NOTE: ELL = English language learners. Formerly ELL = students who passed their state's English-language proficiency examination within the past 2 years. The results for English language learners are based on students who were assessed and cannot be generalized to the total population of such students. Details may not sum to totals because of rounding.
SOURCE: Perie, Grigg, and Donahue (2005).

Exhibit 5. Average Mathematics Scales Scores and Achievement-level Results, by English Language Learners (ELL), Grade 4 Public Schools: By State, 2005

	ELL					Non-ELL					Formerly ELL				
			Percentage of students					Percentage of students					Percentage of students		
State/Jurisdiction	Percentage of all students	Average scale score	Below *Basic*	At or above *Basic*	At or above *Proficient*	Percentage of all students	Average scale score	Below *Basic*	At or above *Basic*	At or above *Proficient*	Percentage of all students	Average scale score	Below *Basic*	At or above *Basic*	At or above *Proficient*
Nation (public)	10	216	46	54	11	89	239	18	82	38	1	240	15	85	35
Alabama	2	‡	‡	‡	‡	98	225	33	67	21	#	‡	‡	‡	‡
Alaska	19	218	47	53	15	81	240	17	83	38	#	‡	‡	‡	‡
Arizona	19	208	60	40	7	81	235	23	77	33	#	‡	‡	‡	‡
Arkansas	3	229	28	72	24	97	236	22	78	34	#	‡	‡	‡	‡
California	31	214	50	50	10	66	238	20	80	36	2	246	8	92	45
Colorado	11	208	58	42	6	88	243	15	85	43	1	‡	‡	‡	‡
Connecticut	4	215	50	50	10	96	243	14	86	44	#	‡	‡	‡	‡
Delaware	4	229	30	70	22	96	240	15	85	37	#	‡	‡	‡	‡
Florida	7	219	43	57	15	90	241	16	84	39	4	230	25	75	21
Georgia	2	208	58	42	4	98	234	23	77	30	#	‡	‡	‡	‡
Hawaii	7	204	64	36	4	93	232	25	75	28	#	‡	‡	‡	‡
Idaho	8	221	37	63	10	92	244	12	88	43	1	‡	‡	‡	‡
Illinois	9	204	64	36	5	91	236	22	78	34	#	‡	‡	‡	‡
Indiana	3	‡	‡	‡	‡	97	240	16	84	39	#	‡	‡	‡	‡
Iowa	4	‡	‡	‡	‡	96	241	14	86	38	#	‡	‡	‡	‡
Kansas	5	‡	‡	‡	‡	94	247	11	89	48	#	‡	‡	‡	‡
Kentucky	1	‡	‡	‡	‡	99	232	25	75	26	#	‡	‡	‡	‡
Louisiana	1	‡	‡	‡	‡	99	230	26	74	24	#	‡	‡	‡	‡
Maine	1	‡	‡	‡	‡	99	241	16	84	39	#	‡	‡	‡	‡
Maryland	3	226	34	66	20	96	239	21	79	39	#	‡	‡	‡	‡
Massachusetts	6	226	32	68	19	93	249	8	92	51	2	‡	‡	‡	‡
Michigan	3	‡	‡	‡	‡	97	238	20	80	38	#	‡	‡	‡	‡
Minnesota	7	222	38	62	14	93	248	10	90	50	#	‡	‡	‡	‡
Mississippi	1	‡	‡	‡	‡	99	227	31	69	19	#	‡	‡	‡	‡
Missouri	2	‡	‡	‡	‡	98	235	21	79	31	#	‡	‡	‡	‡
Montana	3	‡	‡	‡	‡	97	242	13	87	39	#	‡	‡	‡	‡
Nebraska	7	211	56	44	5	92	240	17	83	39	1	‡	‡	‡	‡
Nevada	16	209	59	41	7	84	234	23	77	30	#	‡	‡	‡	‡
New Hampshire	2	‡	‡	‡	‡	98	246	10	90	47	#	‡	‡	‡	‡
New Jersey	3	‡	‡	‡	‡	97	245	14	86	46	#	‡	‡	‡	‡
New Mexico	25	208	58	42	5	75	229	28	72	24	#	‡	‡	‡	‡
New York	5	213	50	50	6	89	240	17	83	38	6	240	15	85	36
North Carolina	6	228	26	74	18	93	242	16	84	41	1	249	10	90	55
North Dakota	1	‡	‡	‡	‡	99	243	11	89	41	#	‡	‡	‡	‡
Ohio	1	‡	‡	‡	‡	99	242	16	84	43	#	‡	‡	‡	‡
Oklahoma	5	222	35	65	11	94	235	20	80	30	#	‡	‡	‡	‡
Oregon	13	215	50	50	12	87	242	15	85	41	#	‡	‡	‡	‡
Pennsylvania	2	218	46	54	17	98	241	17	83	42	#	‡	‡	‡	‡
Rhode Island	6	199	71	29	5	93	236	20	80	32	#	‡	‡	‡	‡
South Carolina	2	‡	‡	‡	‡	98	238	19	81	36	#	‡	‡	‡	‡
South Dakota	4	204	63	37	2	96	243	12	88	42	#	‡	‡	‡	‡
Tennessee	2	‡	‡	‡	‡	98	232	26	74	28	#	‡	‡	‡	‡
Texas	14	226	31	69	15	84	245	10	90	44	2	244	8	92	39
Utah	11	219	42	58	13	89	241	14	86	40	1	‡	‡	‡	‡
Vermont	2	‡	‡	‡	‡	98	243	13	87	43	#	‡	‡	‡	‡
Virginia	7	232	28	72	25	92	241	16	84	40	#	‡	‡	‡	‡
Washington	8	215	46	54	8	92	244	13	87	45	#	‡	‡	‡	‡
West Virginia	#	‡	‡	‡	‡	100	231	25	75	25	#	‡	‡	‡	‡
Wisconsin	6	225	33	67	19	94	242	15	85	42	#	‡	‡	‡	‡
Wyoming	4	223	34	66	15	96	244	12	88	44	#	‡	‡	‡	‡
Other jurisdictions															
District of Columbia	4	206	64	36	7	96	211	55	45	10	#	‡	‡	‡	‡
DoDEA[1]	7	224	32	68	15	93	240	14	86	36	#	‡	‡	‡	‡

#The estimate rounds to zero.
‡Reporting standards not met. Sample size in insufficient to permit a reliable estimate.
[1] Department of Defense Domestic Dependent Elementary and Secondary Schools
NOTE: ELL = English language learners. Formerly ELL = students who passed their state's English-language proficiency examination within the past 2 years. The results for English language learners are based on students who were assessed and cannot be generalized to the total population of such students. Details may not sum to totals because of rounding.
SOURCE: Perie, Grigg, and Dion (2005).

Exhibit 6. Average Mathematics Scales Scores and Achievement-level Results, by English Language Learners (ELL), Grade 8 Public Schools: By State, 2003

	ELL					Non-ELL					Formerly ELL				
			Percentage of students					Percentage of students					Percentage of students		
State/Jurisdiction	Percentage of all students	Average scale score	Below *Basic*	At or above *Basic*	At or above *Proficient*	Percentage of all students	Average scale score	Below *Basic*	At or above *Basic*	At or above *Proficient*	Percentage of all students	Average scale score	Below *Basic*	At or above *Basic*	At or above *Proficient*
Nation (public)	6	244	71	29	6	93	280	30	70	30	1	276	34	66	24
Alabama	1	‡	‡	‡	‡	99	262	47	53	15	#	‡	‡	‡	‡
Alaska	15	260	52	48	11	85	282	27	73	32	#	‡	‡	‡	‡
Arizona	13	245	72	28	5	87	279	31	69	29	1	‡	‡	‡	‡
Arkansas	1	‡	‡	‡	‡	99	272	36	64	22	#	‡	‡	‡	‡
California	20	241	74	26	5	74	275	35	65	26	5	278	33	67	25
Colorado	6	246	71	29	5	94	283	27	73	34	#	‡	‡	‡	‡
Connecticut	3	242	74	26	9	97	282	29	71	35	#	‡	‡	‡	‡
Delaware	3	‡	‡	‡	‡	97	282	27	73	30	#	‡	‡	‡	‡
Florida	5	243	70	30	4	93	276	33	67	27	2	257	52	48	7
Georgia	2	‡	‡	‡	‡	98	273	38	62	23	#	‡	‡	‡	‡
Hawaii	6	229	83	17	3	94	268	42	58	19	#	‡	‡	‡	‡
Idaho	6	254	58	42	7	94	283	25	75	31	#	‡	‡	‡	‡
Illinois	2	249	70	30	8	98	278	31	69	29	#	‡	‡	‡	‡
Indiana	2	‡	‡	‡	‡	98	282	25	75	31	#	‡	‡	‡	‡
Iowa	2	‡	‡	‡	‡	98	285	24	76	34	#	‡	‡	‡	‡
Kansas	3	251	67	33	3	97	285	22	78	35	#	‡	‡	‡	‡
Kentucky	1	‡	‡	‡	‡	99	274	35	65	23	#	‡	‡	‡	‡
Louisiana	1	‡	‡	‡	‡	99	268	41	59	16	#	‡	‡	‡	‡
Maine	1	‡	‡	‡	‡	99	281	26	74	30	#	‡	‡	‡	‡
Maryland	2	‡	‡	‡	‡	98	278	33	67	30	#	‡	‡	‡	‡
Massachusetts	2	242	73	27	8	97	293	19	81	44	1	255	60	40	11
Michigan	2	‡	‡	‡	‡	98	278	32	68	30	#	‡	‡	‡	‡
Minnesota	6	‡	‡	‡	‡	93	292	19	81	45	#	‡	‡	‡	‡
Mississippi	1	‡	‡	‡	‡	99	263	48	52	14	#	‡	‡	‡	‡
Missouri	1	‡	‡	‡	‡	99	277	32	68	26	#	‡	‡	‡	‡
Montana	4	243	73	27	3	96	288	18	82	37	#	‡	‡	‡	‡
Nebraska	3	242	78	22	2	97	285	23	77	36	#	‡	‡	‡	‡
Nevada	9	236	79	21	4	90	273	36	64	23	1	‡	‡	‡	‡
New Hampshire	1	‡	‡	‡	‡	99	286	22	78	35	#	‡	‡	‡	‡
New Jersey	1	‡	‡	‡	‡	98	284	25	75	36	1	‡	‡	‡	‡
New Mexico	16	239	77	23	2	84	268	42	58	16	#	‡	‡	‡	‡
New York	4	237	77	23	4	87	282	28	72	32	9	278	33	67	27
North Carolina	3	252	58	42	8	96	283	27	73	33	1	‡	‡	‡	‡
North Dakota	1	‡	‡	‡	‡	99	287	19	81	35	#	‡	‡	‡	‡
Ohio	1	‡	‡	‡	‡	99	284	25	75	33	#	‡	‡	‡	‡
Oklahoma	4	252	60	40	12	96	272	36	64	21	#	‡	‡	‡	‡
Oregon	7	253	60	40	10	93	285	25	75	35	#	‡	‡	‡	‡
Pennsylvania	1	‡	‡	‡	‡	99	281	27	73	31	#	‡	‡	‡	‡
Rhode Island	4	224	89	11	1	96	274	34	66	24	#	‡	‡	‡	‡
South Carolina	1	‡	‡	‡	‡	99	282	28	72	30	#	‡	‡	‡	‡
South Dakota	2	‡	‡	‡	‡	98	288	19	81	37	#	‡	‡	‡	‡
Tennessee	1	‡	‡	‡	‡	99	271	39	61	21	#	‡	‡	‡	‡
Texas	6	242	74	26	3	92	284	25	75	33	1	276	29	71	20
Utah	6	249	63	37	8	93	281	26	74	31	1	‡	‡	‡	‡
Vermont	1	‡	‡	‡	‡	99	288	22	78	38	#	‡	‡	‡	‡
Virginia	4	260	49	51	13	96	285	24	76	34	#	‡	‡	‡	‡
Washington	4	249	68	32	11	96	287	23	77	37	#	‡	‡	‡	‡
West Virginia	#	‡	‡	‡	‡	100	269	40	60	18	#	‡	‡	‡	‡
Wisconsin	3	269	44	56	19	97	285	23	77	36	#	‡	‡	‡	‡
Wyoming	4	251	61	39	3	96	283	22	78	30	#	‡	‡	‡	‡
Other jurisdictions															
District of Columbia	3	‡	‡	‡	‡	97	246	69	31	7	#	‡	‡	‡	‡
DoDEA[1]	4	260	54	46	10	96	285	23	77	33	#	‡	‡	‡	‡

#The estimate rounds to zero.
‡Reporting standards not met. Sample size in insufficient to permit a reliable estimate.
[1] Department of Defense Domestic Dependent Elementary and Secondary Schools
NOTE: ELL = English language learners. Formerly ELL = students who passed their state's English-language proficiency examination within the past 2 years. The results for English language learners are based on students who were assessed and cannot be generalized to the total population of such students. Details may not sum to totals because of rounding.
SOURCE: Perie, Grigg, and Dion (2005).

NCLB Provisions for ELL Inclusion and Assessment

Under NCLB, ELL students are defined as being an individual:

A). who is aged 3 through 21;

B). who is enrolled or preparing to enroll in an elementary school or secondary school and who:

(i) was not born in the United States or whose native language is a language other than English;

(ii) is a Native American or Alaska Native, or a native resident of the outlying areas;

(iii) comes from an environment where a language other than English has had a significant impact on the individual's level of English language proficiency; or

(iv) is migratory, whose native language is a language other than English, and who comes from an environment where a language other than English is dominant; and

C). whose difficulties in speaking, reading, writing, or understanding the English language may be sufficient to deny the individual the:

(i) ability to meet the state's proficient level of achievement on state assessments described in section 1111(b)(3);

(ii) ability to successfully achieve in classrooms where the language of instruction is English; or

(iii) opportunity to participate fully in society. [Title IX, Part A—Definitions, Sec. 9101 (25)]

This definition has much flexibility at the state level. The U.S. Department of Education (2004) states:

> *LEP is not a demographic group per se, but a classification that changes as a student gains language proficiency. Its membership can change from year to year with language proficient students exiting each year and new LEP students entering each year. Since LEP students exit the subgroup once they attain English language proficiency, states may have difficulty demonstrating improvements on state assessments for this student subgroup.*
>
> and
>
> *The NCLB definition of a limited English proficient student gives states flexibility in defining the students who constitute the LEP subgroup. For example, a state has the flexibility to define narrowly the LEP subgroup as only those students receiving direct, daily LEP services. A state could also define the group more broadly to include both students receiving direct services and students being monitored based on their achievement on academic assessments.*

A key point in the foregoing is the notion that the U.S. Department of Education does not equate the term "LEP" with a stable demographic group, a notion that also resonates with states' own historical characterizations of the group for educational policy and accountability purposes. NCLB, however, does not allow states completely free reign in identifying and tracking ELL students. Under prior instantiations of the ESEA and requirements for support provided to states under the legislation, states historically have

been expected to identify ELL status upon students' entry into schools drawing on three sources: 1) evidence of a non-English language spoken at home based on a home language survey; 2) performance on an English language proficiency test; and 3) teacher-school personnel clinical judgment of readiness for English academic work, which can include analysis of performance based on locally developed assessments and student interviews. Drawing on such evidence, states can stipulate procedures used for identification of ELL students in their unique state education codes.

NCLB Title III requires that states have English language development standards in place for ELL students that cover reading, writing, listening, and speaking, and administer language proficiency assessments based on these standards. Titles I and III require that states adopt English language proficiency tests covering reading, writing, listening, and speaking, and that these assessments be administered annually in grades K-12 to all ELL students.[3] Further, under NCLB Title III, states are directed to adopt Annual Measurable Achievement Objectives (AMAOs) that set goals for ELL students' aggregate progress in acquiring English proficiency and in attainment of English language proficiency. AMAOs are monitored annually at the school district level and for the state as a whole. NCLB requires that states submit a biennial report on their progress as a funding requirement for Title III support.

Under NCLB, states must include all ELL students in their assessments, and all ELL students are held to the same state academic learning standards and adequate yearly progress goals as other students and major subgroups of students in reading and mathematics. Beginning in 2007-08, these policies will apply in science as well. States, however, have the option of exempting ELL students from English language testing in these areas for up to 3 years following initial entry into the schools. After their third year in the school, ELL students must be tested in English. During the interim period, states are encouraged to assess students in their primary language in the target subject matter areas to the extent primary language assessments are available and are aligned with state academic learning standards. In addition, during the interim period, states may administer accommodated versions of English language assessments of reading and mathematics (and science assessments starting in 2007-08) provided that the accommodated assessments do not alter the measurement objectives of the English version assessments.

The implementation of NCLB assessment provisions for ELL students has been a developmental process in which states and the U.S. Department of Education have negotiated interpretation of NCLB requirements. One of the first clarifications that emerged was that states would not be required to have 100 percent of (formerly) ELL students attain full English proficiency by 2013-14. This clarification was in large part an acknowledgement that ELL students can enter U.S. schools at any point in the K-12 range, and that research and practical experience indicate that full acquisition of English proficiency can take several years (Hakuta, Butler, and Witt, 2000).

In 2004, the U.S. Department of Education (February 2004) issued formal guidance allowing greater flexibility in the assessment of new immigrant ELL students under NCLB. While states are still required to administer an English language proficiency test to all ELL students on an annual basis from the outset of students' entry into the schools, states were granted permission to test ELL students' reading skills in English in their first year in schools, thus allowing assessment of students from language backgrounds where no native

[3] NCLB Title III also requires that states report a derived English language "comprehension" score based on performance on the adopted English language proficiency testing system.

language reading achievement test was available. Use of the state mandated mathematics assessment in English accompanied by accommodations was also endorsed as a remedy. In addition, the Department of Education's February 2004 guidance allows new immigrant ELL students to count toward meeting the 95 percent inclusion requirement for a subgroup in a state assessment, while allowing a state the option of not including the test results of these new ELL immigrant students in the calculation of AYP.

Another major adjustment in NCLB provisions for ELL students was that states would be permitted to include the scores of *former* ELL students in the computation of AYP gains for up to 2 years after leaving ELL status. This adjustment is intended to help states better establish evidence that their programs and services for ELL students are indeed helping students achieve and sustain progress toward AYP targets.

Key NCLB Factors Affecting Interpretation of ELL Scores on State Assessments and the States' Role

The foregoing is a synopsis of NCLB provisions for states' assessment of ELL students that are central factors affecting the validity of inferences made about these students based on their participation in a state assessment system. These factors include states' enactments of:

- criteria and procedures for identification of ELL students and for exiting from ELL status;
- implementation of English language proficiency assessments required annually for ELL students under NCLB;
- inclusion policies for ELL students on state assessments in English; and
- provisions of alternate language assessments and assessment accommodations.

Awareness of these factors and the challenges they present to establishing the validity of state assessments for ELL students under NCLB have been examined by a number of investigators (see Abedi 2004b, for example).

In addition, Abedi (2004b) points out that individual states and school districts vary in the ELL subpopulations they serve, their relative density in a state, and across school districts. He notes that local school districts and schools with a high density of ELL students are most at risk for failing to make AYP gains, and that this possibility becomes more likely for these schools as AYP goals are raised.

It seems clear that the forgoing factors are substantive issues for individual states under NCLB even without immediate concern for the comparability of states' progress and results in implementing NCLB *across states*. It may be argued that this concern for *within state* resolution of NCLB goals and implementation is crucial given the central role played by states in local educational governance, and given variations in the character, size, and distribution of ELL populations within states. Indeed, it may be argued that establishing the validity of state accountability systems under NCLB is central toward achieving the goals of NCLB (Marion et al., 2002; see also Fuhrman and Elmore, 2004). Arguably, this should include components of state accountability for ELL student achievement attuned to the unique characteristics of a state and its capacity to martial resources to address issues such as those cited above.

States Progress in Implementing NCLB ELL Assessment Provisions

The foregoing emphasis on attention to within state resolution of NCLB issues for ELL students is echoed in the U.S. Department of Education (2005) report to Congress on states' progress in implementing NCLB Title III provisions for ELL students during 2002-04. This report argues that Congress and educational stakeholders need to avoid direct, simple comparison of NCLB compliance data across states due to legitimate, but idiosyncratic implementation of NCLB ELL inclusion and assessment policies by individual states. Despite idiosyncrasies among states, the 2005 report indicates that states have made much progress in implementing the ELL assessment provisions of NCLB. For example, as the report indicates, prior to enactment of NCLB in 2002, only 7 states had developed English language proficiency standards, whereas by 2004 all 50 states (and the District of Columbia and Puerto Rico) had adopted such standards that are linked conceptually to state academic standards. Further, all states have submitted evidence to the U.S. Department of Education that they have adopted or are on the way to adopting English language proficiency tests aligned with English development standards.

In late fall 2004, the U.S. Department of Education initiated a formal peer evaluation of states' NCLB implementation that includes attention to ELL provisions of the law. The letter to state chief school officers (U.S. Department of Education, October 29, 2004) states:

> *The peer review process will not directly examine a State's academic standards, assessment instruments, or specific test items. Rather, it will examine evidence compiled and submitted by each State that is intended to show that the assessment system as implemented meets NCLB requirements. Such evidence may include, but is not limited to, results from alignment studies; results from validation studies; written policies on providing accommodations for students with disabilities and LEP students; written policies on native-language testing of LEP students (if applicable); and score reports showing disaggregation of student achievement data by the statutorily specified student subgroups. Sufficient evidence must be provided to convince an experienced professional that the assessment system can be implemented in a manner that meets NCLB requirements.*

These reviews will take substantial time to complete and the results will represent the next important, major set of benchmarks regarding the implementation of assessment provisions of NCLB for ELL students.

NCLB ELL Progress in Four States

As mentioned earlier, attention is given in this report to the four states of California, New York, Texas, and Washington and to issues that arise during their implementation of NCLB provisions for ELL students and in considering how NAEP might be of value in interpreting the performance of ELL students in these states. In 2001-02, California (with 1,512,655 ELL students) was home to 32 percent of the nation's 4,747,763 ELL students in grades K-12. ELL students in California constituted 24 percent of all California students in this grade range. In 2001-02, New York had 266,774 ELL students in grades K-12 who constituted another 6 percent of the nation's total ELL population and 9 percent of all students in New York. In 2001-02, Texas enrolled 601,791 ELL students—13 percent of the nations' ELL students and 15 percent of the state's students. Washington enrolled 70,431 ELL students in

2001-02—about 1 percent of all ELL students nationally, and about 7 percent of all Washington students in grades K-12.

The four states under consideration vary in their geographic location, ELL student populations, state education system governance, and emerging response to NCLB provisions. Consideration of some of these characteristics are valuable for better framing issues faced by states in interpreting ELL NAEP results and in better understanding challenges in exploring how NAEP may or may not be useful for cross-state comparisons of ELL performance on state tests and NAEP.

The U.S. Department of Education (2005) biennial evaluation report to Congress on the implementation of NCLB Title III provisions indicates the four states (and all states as a whole) have made substantial progress in the implementation of NCLB—with the caveat cited above that each state has developed its own interpretation of NCLB mandates given state education code and policies.

While each state has implemented annual English language proficiency assessments, none of these four states is using the same English language proficiency test: California is relying on the California English Development Test (CELDT); New York utilizes the New York State English Second Language Test (NYSESLAT) and Language Assessment Battery (LAB-R); Texas uses the Texas Language Proficiency Assessment System (TELPAS—Reading Proficiency Test in English and Observation Protocol); and Washington relies on the Language Assessment Scales (LAS) and Washington Language Proficiency Test (WLPT). Each state, through its own standards-setting process, determines how student performance on these assessments is used as evidence for initial identification of English language learner status and for evidence of sufficient English proficiency to exit ELL status. Further, each state determines how performance on the state's English language proficiency assessments is weighed with other evidence of English language proficiency in making a decision to exit students from ELL status. Additional indicators may include students' performance on state subject matter tests in English language arts, teachers' professional judgments of ELL students' readiness to benefit from English instruction, and parents' advice on students readiness to benefit from instruction in English. To add to the complexity, local school districts in some states such as California have the final say on how to weigh criteria, thereby leading to possible variation within states in ELL exit decisions.

At present, there are no extant empirical validity studies of similarities and differences of performances on the aforementioned language proficiency assessments, let alone studies of commonalties and differences in standard-setting procedures and decision criteria for ELL identification and exit. Another complication is the fact that the four states are at different points in finalization of their English language proficiency assessments and English language development metrics. For example, California has just completed development of a common vertical scale for the CELDT English proficiency assessment for the entire K-12 sequence, while Texas has yet to complete implementation of the observation protocol that will complement its RPTE English proficiency test; and Washington is just completing selection of a new English language proficiency test that will be aligned with their English language development standards. Under NCLB Title III, the final selection and implementation of English language proficiency assessments is made all the more complicated because they also have to be used to measure ELL students' aggregate growth over time in acquiring English and attainment of English proficiency—hence for example, California's implementation of a vertical scale of English language proficiency on their CELDT assessment.

While evaluated by U.S. Department of Education as a means toward compliance with the subgroup inclusion provisions of NCLB, the four states of California, New York, Texas, and Washington differ in how they include students in their assessment systems. Not all of the states have annual tests in reading and mathematics aligned with state standards for grades 3 through 8. California (California Standards Test—CST) and Texas (Texas Assessment of Knowledge and Skills—TAKS) have such assessments in place, but New York and Washington are in the process of adopting such assessments over the next 2 years. There is an interaction between inclusion in accountability testing and the availability of native language assessments provided by states.

In California, ELL students are required to take the CST reading and mathematics assessments in English (along with the norm referenced California Achievement Test-6) after one year in California schools, with acceptable accommodations as necessary. In the first year, California ELL students may be tested in a non-English language, provided tests are available and approved. Under Texas education code for ELL services and accountability, Texas offers a Spanish version of TAKS in reading and mathematics to ELL students in grades 3-6 who receive Spanish-English bilingual instruction. (The state also offers a Spanish version of TAKS in grade 4 writing and grade 5 science to these students.) New York education code allows students classified as ELL and provided bilingual instruction to be tested in one of six non-English primary languages (Chinese, Haitian Creole, Korean, Russian, and Spanish) for which assessments have been developed at the fourth-, fifth-, and eighth-grade levels in mathematics and social studies. Washington requires that all ELL students take the mandated WASL tests in reading, writing, mathematics, and science at grades 4-10, with a new WASL science test at grade 5 added in 2003-04. All of the four states permit similar assessment accommodations for ELL students taking the state mandated NCLB mathematics tests.

New York and Texas have been permitted by the U.S. Department of Education to use one of their English language proficiency assessments as an acceptable alternate for ELL students in reading. ELL students scores on these tests count toward participation in the NCLB mandated reading test, and scores earned on these proficiency tests are used as AYP indicators in reading for these ELL students.

Challenges and Opportunities in Comparing NAEP and State Assessment Performance for ELL Students

The staff on this study conducted interviews with NAEP state coordinators for California, New York, Texas, and Washington supplemented by interviews with key assessment and accountability staff and officers for three of these states.[4] The purpose of the interviews was to explore prospects for examining relationships between ELL student performance on NCLB mandated state tests and performance on NAEP. The following summary of interview findings is organized by themes in the responses that cut across issues raised by the interview protocol questions. The themes pursued in these interviews are found in Exhibit 7. In reviewing the summary of the findings, it is essential that they be

[4] The persons interviewed included Eric Zilbert (California State Department of Education, State NAEP Coordinator), John Smith (Texas Education Agency, State NAEP Coordinator), Laura Ayala (Texas Education Agency, Director of Limited English Proficient Assessments), David Moore (New York State Department of Education, State NAEP Coordinator), David Abrams (New York State Department of Education, Assistant Commissioner Office of Standards, Assessment and Reporting), Pete Bylsma (Washington State Office of the Superintendent of Public Instruction, Director Research and Evaluation, Assessment and Research Section), and Kathryn Sprigg (Washington State Office of the Superintendent of Public Instruction, State NAEP Coordinator).

treated as anecdotal and not generalizable. The goal of the interviews was to probe consciousness of issues so that the NAEP program might better understand its options in developing strategies promoting more informed use of NAEP test scores for ELL students, given state's response to NCLB.

Exhibit 7. Question Areas Guiding Interviews of State NAEP Directors and State Assessment and Accountability Staff

1) What is the status of your state ELL student assessment and accountability proposal to the U.S. Department of Education with respect to NCLB requirements? Which issues are resolved and which remain to be resolved? What steps and timelines are in place to resolve issues? Who else should be contacted?

2) How does your state determine ELL (LEP) status under NCLB and under NAEP guidelines?

3) Is the population of ELL students participating in your state achievement tests, compliant with NCLB, arguably the same as the population identified as ELL by the sample participating in the NAEP state assessment? How does it differ and how might a reconciliation be made?

4) ELL students present on an assessment day have three options for participation in state assessments and NAEP: 1) be included without accommodations; 2) be included with accommodations; and 3) be exempted from the assessment. Given the assessment procedures in place for both types of assessments, what similarities and differences might arise in terms of the classification rates of ELLs into the 3 categories across the state assessments and NAEP?

5) Specific to NAEP, what do we know about trends in ELL participation in NAEP in your state?

6) From your perspective, how is your state taking on the study of relationships between ELL scores on state assessments tied to NCLB and NAEP? What obvious issues arise in thinking about comparing ELA and Math scores on state assessments and NAEP?

7) Is there a plan? What are the components of this plan and what is a possible schedule for implementation of the plan?

8) What challenges and opportunities do you perceive in attempting to study relationships among ELL students in state assessments and NAEP across states?

9) What are your thoughts on the use of a common English language proficiency test among states as a tool for examining the performance of ELL students on state tests and NAEP across states?

10) Other ideas on relevant issues tied to the scope of work?

Initially, in the conduct of this study, no interviewee was aware of a formal plan in their state for comparing or analyzing relationships in-depth between NAEP scores and state achievement test scores for ELL students in reading, mathematics, or science. This orientation changed over the course of the study. Initially, all interviewees expressed an interest in such research with the proviso that such research was not the most immediate and pressing priority given their responsibilities to the NAEP program or their state assessment and accountability offices. The interviewees from all four states judged that there were limited incentives and resources at present for their states to investigate in detail relationships between the performance of ELL students on NAEP reading and math

assessments and the performance of these students on state reading and math tests allied with NCLB. Interestingly, by the end of study, evidence began to appear that states were beginning to develop formal plans for studying association between NAEP scores and state achievement and English proficiency scores. For example, Washington has begun to conduct exploratory analysis of relationships between NAEP scores and state test scores for ELL students, and it is requiring the testing firm implementing its newly contracted English proficiency test to examine relationships between ELL students' NAEP test scores, state English language proficiency test scores, and state achievement test scores. California also will continue to pursue studies of the relationship between NAEP scores and state achievement test scores for ELL students. Results from the California research are mentioned later in this report.

While the incentive to investigate relationships between NAEP and state ELL achievement data in detail is influenced by the absence of legal or policy mandates to do so, there are also a number of policy, accountability, and data quality factors that make detailed study of relationships between NAEP performance and state test performance problematic. These factors raise issues that need resolution as part of a strategy to propose studying relationships.

The interviewees commented that while NAEP is a valuable indicator of student achievement, it is not necessarily well-aligned with state standards and state assessments intended to assess student's attainment of state standards. Despite this reservation, the four state NAEP coordinators did believe that increases in state NAEP scores coupled with increases in state achievement scores would be arguable evidence of improvements in ELL achievement in their states. The state NAEP coordinators commented that a fundamental challenge they faced was clearly communicating the meaning of NAEP results in their state to educators and the public and answering such questions as what students can and cannot do in classrooms at various NAEP proficiency levels. All the NAEP coordinators interviewed indicated that, in their roles as state NAEP coordinators, they were separated from state accountability and assessment units that were responsible for monitoring the development of state NCLB plans and the monitoring of compliance with the law. The coordinators commented that an understanding of the performance of ELL students and other students on state achievement tests needs to build on the specific assessment frameworks and policy provisions guiding the implementation of the state assessment system, as well as on the policies and practices for inclusion of ELL students in state assessments.

Two of the non-NAEP coordinators interviewed echoed these concerns and expressed doubts about the utility of NAEP scores for informing the analysis of ELL achievement improvements in their states. States have specific accountability and assessment strategies linked to subject matter learning standards, state curriculum improvements, and teacher staff development programs. NAEP does not have these concrete connections to their state. Further, ELL students are likely to be low-achieving students, and state tests under NCLB may be more sensitive to their improvements in achievement, while NAEP is more sensitive to improvements in the higher ability levels.

The coordinators interviewed were well aware of NAEP policies for inclusion and accommodations of ELL students in their state, but had not given concerted attention to whether the identification rates for ELL students and inclusion rates for ELL students with and without accommodations on NAEP could be reconciled with similar data for their state assessments in reading and mathematics. All the coordinators acknowledged the need for

more study of relationships between ELL student NAEP scores and state achievement test scores.

During the interviews, the four state NAEP coordinators were shown data on inclusion of ELL (non-special education) students and accommodation rates in recent NAEP assessments. The state coordinators did not comment at length on the inclusion/ accommodation trends for their state. They did note the importance of local NAEP administrator decisions in determining whether and how ELL students participate (including accommodations decisions), while at the same time enforcing the NAEP guidelines for inclusion of ELL students.

The non-NAEP coordinators suggested that the rates of ELL exclusions from regular English language achievement testing in their states are a reflection of state education code requirements for testing ELL students and for including these students in the state assessment system. NAEP exclusion rates for ELL students are also a reflection of these state policies and practices. As these policies and practices change, NAEP participation rates for ELL students would be expected to change as long as NAEP bases its identification of ELL students upon current state regulations and practices for identification and assessment of ELL students.

All persons interviewed commented that inclusion and accommodation policies within a state are not necessarily reflected accurately in local school assessment practices for ELL students. Several interviewees also noted that state data reporting and tracking systems are far from perfect for NCLB purposes, and they pointed out that the refinement of these systems and their accuracy is a developmental process that requires adequate systems design, resources to implement systems, and improvements in the capabilities of local school districts and schools to adopt and implement these systems with fidelity.

Results from a California Study of ELL Performance on NAEP and State Assessment

A recent study in California by the state NAEP coordinator suggests that, under the right circumstances, ELL gains on a state test might be associated with gains on NAEP (Zilbert, 2004)—see Exhibits 8a and 8b. Mean NAEP reading scores from Los Angeles for all fourth graders participating in the state NAEP in 2002 and the urban trial NAEP in 2003 showed a modest, but not reliably estimable gain (effect size .08). Overall student performance for Los Angeles on the CST in English Language Arts also showed a slight gain in means from 2002 to 2003, but this effect was again not reliably estimable (effect size .16). In contrast, the Los Angeles data for ELL students show notable evidence of gains in mean reading scores on the NAEP scale and in mean English Language Arts scores on the CST from 2002 to 2003. The NAEP reading means for ELL students showed a gain in 2003 from 2002 corresponding to an effect size of .27—a notable gain on the NAEP reading metric. The gain for ELL students in mean scores on the Language Arts component of the California Standards Test from 2002 to 2003 showed a gain corresponding to an effect size of .28.

Exhibit 8a. NAEP and CST Scores for All Grade 4 Students in the Los Angeles Unified School District (LAUSD), 2002 and 2003

LAUSD Grade 4	NAEP Reading		CST in English Language Arts	
	Avg. Score	Std. Deviation	Avg. Score	Std. Deviation
2003	194	37	326.1	45.4
2002	191	37	318.9	46.3
Change	3	Avg. 37	7.2	Avg. 45.85
Effect Size	*.08*		*.16*	

NOTE: Exhibit 8a shows the results of the same analysis applied to the 2002 and 2003 NAEP and CST results for the LAUSD. The NAEP estimate is somewhat lower than that for the state assessment, and both measures indicated limited progress.
SOURCE: Zilbert (2004).

Exhibit 8b. NAEP and CST Scores for Grade 4 ELL Students in the Los Angeles Unified School District (LAUSD), 2002 and 2003

	NAEP Reading	CST in English Language Arts
2003	183	310
2002	173	297
Change	10	13
Effect Size	*.27*	*.28*

NOTE: LAUSD English learners in the NAEP sample showed a statistically significant improvement in NAEP reading scores from 2002 to 2003. Analysis of the CST data for English learners confirms this finding. Effect size computations in Exhibit 8b are based on standard deviation and average standard deviation statistics reported in Exhibit 8a for all LAUSD students.
SOURCE: Zilbert (2004).

Data such as the foregoing are provocative and invite follow-up analyses. Are the gains for ELL students being affected by changes in the density of ELL students in the Los Angeles School system? ELL students comprised 43 percent of Los Angeles fourth graders in 2002, but then jumped to 54 percent of Los Angeles fourth graders in 2003. What dynamics were at play here, and could these dynamics have been associated with the gains? The stereotypical belief that increases in the number of ELL students at a grade level in a school district would lead to lower test performance in assessments at that grade level was contradicted. It also is hard to believe that an additional year of experience in implementing school reform could lead to such a significant improvement in ELL students' scores, but then again, is there evidence of significant changes in instruction and teacher support in one year's time that could be associated with the gains? What will 2005 NAEP mean reading performance and mean performance on the CST Language Arts Test look like in Los Angeles (and statewide) for ELL and non-ELL students? Will findings such as those encountered in Los Angeles be mirrored in other states and local school district jurisdictions?

National Implications and Future Directions for NAEP and State Collaborations

The findings of this report suggest the possibility that NCLB is exerting both immediate and long-range impact on states' ELL assessment and accountability policies and practices, and that these impacts may affect the utility of NAEP as an indicator of ELL students' achievement performance within a state. This point needs emphasis as a major conclusion about the validity of state test scores drawn from the present study. Validity most generally concerns the logical and empirical defensibility of inferences drawn from test scores (Messick, 1989). As acknowledged by the National Research Council (Koenig and Bachman,

2004) and the U.S. Department of Education (2005), the population category "English language learner" (or "limited English proficient") is operationalized uniquely by every state. In general terms, this population category refers to students for whom English is a second language and who are evaluated to need further development of English language skills in order to be instructed effectively in English. However, under NCLB terms, adequate yearly progress of ELL students on state achievement tests can only be understood in terms of how this population is defined by a state and how a state operationalizes exit from ELL status. For this reason, if NAEP is to contribute to understanding state achievement of ELL students, it needs to reference the particular ELL population defined by a state.

The NAEP state coordinators and other state assessment personnel interviewed for this project were not sanguine that cross-state comparisons were worthwhile to their state accountability systems and noted that attempting such comparisons was of limited defensibility for the reasons cited previously. Further, they were not convinced that adoption of a common English language proficiency assessment across states was a practical way to begin to explore cross-state comparisons given their own state's trajectory and processes for adopting English language development standards and identifying appropriate English language proficiency tests.

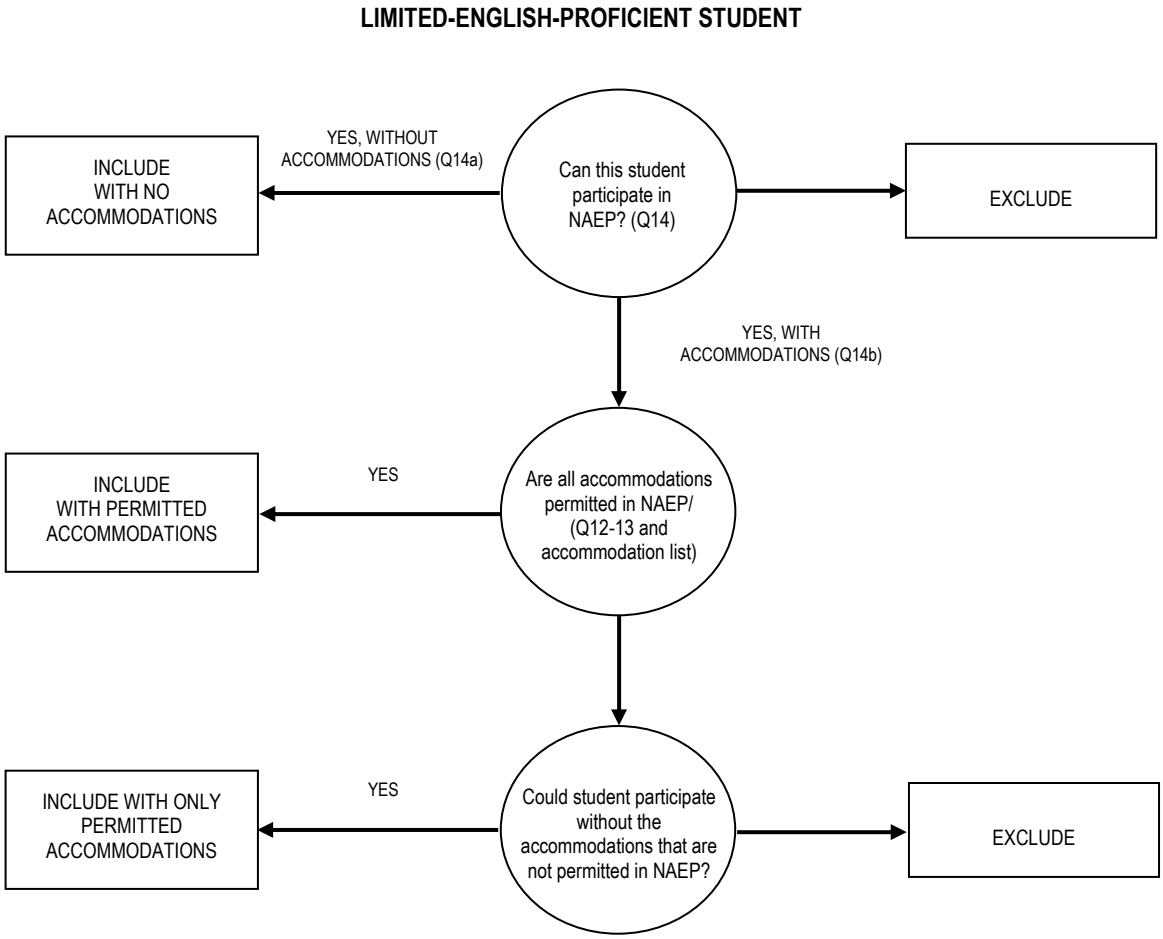
Turning to other relevant perspectives on cross-state comparisons of ELL data, NCES has a long tradition of establishing the highest quality standards for reporting educational statistics nationally. NCES' attention to NAEP reporting has focused primarily on accuracy and validity in aggregate statistics tied to key reporting variables that show stable population definitions across states and urban jurisdictions.

Over the past several years, the National Research Council (NRC) has collaborated with NCES to sponsor a series of workshops and seminars addressing the quality of data reporting on large scale assessments for students with disabilities and English learners. The most recent outcome of this effort is the *Keeping Score for All Report* (Koenig and Bachman, 2004). The contents of this report (which overviews earlier NRC and NCES efforts and newly commissioned studies) make clear that variations in state policies and practices for inclusion and accommodation of ELL students (and students with disabilities) can introduce error into statistics reporting on the test performance of these subgroups, and that this error affects not only the accuracy of reporting population parameters for these subgroups, but also the accuracy of reporting population parameters for the entire population as a whole and for other partitioning of subgroups from the population with ELL students extracted. *The Keeping Score for All Report* thus suggests that better alignment of inclusion and accommodation policies between state assessments and NAEP is critical for purposes of cross-national comparisons of subpopulation results across states.

A fundamental question concerns the manner in which NAEP makes decisions about how to include ELL students and whether to offer them accommodations. This must be considered in light of each state's own policies and practices for including and accommodating ELL students in their state assessments. Exhibit 9 displays the decision tree implemented by the NAEP program in 2005 based on the NAEP background questionnaire for students with disabilities and English language learners. This branching tree outlines the steps used by site administrators as described earlier in this report in deciding whether to include ELL students in NAEP, and if included, whether to allow assessment accommodations permitted by NAEP. In addition, in the 2005 assessment, NAEP field tested a second decision tree devised by the National Assessment Governing Board (NAGB). This decision tree was designed to tie ELL student participation and possible

accommodation in NAEP directly to state policies and practices for inclusion of ELL students in state assessments. This field test has proven workable and has led to changes.

Exhibit 9. NAEP 2005 Decision Tree for Inclusion and Accommodation of ELLs



Implementation of the new decision tree will improve interpretation of NAEP results for ELL students within states. However, implementation of the new decision tree would allow uncommon definitions of ELL students and NAEP inclusion decisions *across* states, thereby interfering with the goal of having NAEP educational statistics provide whole population estimates. Implementation of the new procedure will putatively help make national comparisons ELL performances on NAEP better grounded in the policy sense. Trying to force a common definition of ELL students and assessment inclusion based on uniform standards for inclusion and accommodations (subject to some remaining NAEP caveats on acceptable accommodations) would distort the meaning of NAEP ELL data in the policy sense, simply because this data would not be as applicable to the reality of state accountability and assessments under NCLB.

Implementation of these decision rules and any variants for future NAEP administrations will need careful study, given past experience that states and local jurisdictions vary in their actual inclusion and accommodation practices for ELL students (and students with disabilities). Special attention would need to be given in studies of this implementation to address whether following NAEP inclusion and accommodation

procedures for ELL students would change the composition of the ELL student population maintained by a state as part of its accountability under NCLB. It has been suggested by NAGB that a better procedure would be to have NAEP adhere as closely as possible to states' own policies and procedures for inclusion and assessment accommodation for ELL students. This would help states to better interpret NAEP ELL results in light of state accountability and assessment policies developed under NCLB. Studies of the decision-making practices followed at the local level by NAEP site administrators in implementing ELL inclusion and accommodation practices are needed in order to understand challenges and anomalies faced by site administrators in making grounded and consistent decisions within and between NAEP administration sites. Although NCES is implementing such studies (Goldstein, 2005), more will be needed in order to clarify whether changes in state ELL policies under NCLB will lead to stable policies and practices for inclusion in NAEP across biennial cycles of NAEP administration.

Other research is needed on the relationship between ELL students, state assessments, and NAEP in order to improve the valid use of NAEP in helping to interpret the meaning of state assessment scores for ELL students. As cited earlier in this report, ELL student populations and state policies and practices for accountability and assessment are heterogeneous from state to state for a number of reasons that bear on a deeper range of issues. With regard to English language proficiency assessment, it is unrealistic under NCLB and state education policies to expect states to adopt common English language proficiency at this point in time. However, as Abedi (2004a) points out, there are several active state consortia that are at various phases in the development of common English proficiency assessments even as the participating states use interim instrument to meet NCLB guidelines. Research on the progress of states adopting and implementing common English proficiency tests and common guidelines for inclusion of ELL students in state assessments based on results from these assessments conceivably would allow for improved comparisons of NAEP results for ELL students across states.

Other research should examine more carefully the relationships between ELL students' state assessment performances and NAEP performances. In addition to studies of relationships between state and NAEP assessment performances based on state definitions of ELL status and criteria for exit from ELL status, research needs to be conducted on a systematic basis to determine how specific factors are related to NAEP and state assessment performance of ELL students (e.g., factors that are specific to a non-English background, length of exposure to U.S. schooling in English, the relationship of socioeconomic status to ELL status, length of exposure and types of exposure to English language, and first language development programs). While NAEP should contribute to such research, it will be incumbent on states to assume considerable responsibility for such studies given the utility of this research for better understanding their idiosyncratic implementation of NCLB and the state accountability and assessment practices for ELL students under their jurisdiction.

In closing, the results of the present project suggest that it would be useful for NAGB and NCES to help states better understand strategies they might adopt to make NAEP data for ELL students more useful, given the need for states to comply with NCLB. It would be very helpful if the NCLB offices of the Department of Education were brought into the development of these strategies. States also need to be consulted as part of the development of such strategies, given the states' own policy and accountability systems and their needs to improve educational outcomes for ELL students. Among the possibilities to consider as a next step are a series of meetings and workshops to explore options and possibilities for collaborations among states. Such collaborations could share data or explore

methods for studying relationships between state ELL assessment data and NAEP ELL data. Other possibilities include exploration of issues associated with understanding the relationships between ELL data and data of other groups tied to NCLB reporting requirements. Timing of these pursuits is important. In mid 2006, the U.S. Department of Education (July 2006) announced a "LEP Partnership" initiative that represents an important stride in helping states improve assessment of English language learners pursuant to NCLB. The Partnership was founded by the Department's Office of English Acquisition in collaboration with the Council of Chief State School Officers, the Comprehensive Center on Assessment and Accountability, the National Council of La Raza, and the Mexican American Legal Defense and Educational Fund. The Partnership will:

- Help states measure what LEP students know and what they have yet to learn in all subjects so instructional decisions can be based on valid and reliable data;
- Provide technical assistance and support to states to allow them to continue their ongoing development of valid and reliable assessments; and
- Identify best practices in providing accommodations to LEP students that do not compromise accuracy or academic achievement.

The Department is disseminating commissioned reports and materials supporting these objectives, including related instructional objectives. A first series of reports was made available in fall 2006 by the U.S. Department of Education (October 2006).

A logical next step would be for the Department to convene state NAEP coordinators in conjunction with state LEP Partnership members to explore systematically how NAEP assessments might inform states' analyses of progress toward NCLB achievement goals for ELL students. NAEP coordinators already have become aware of the importance of collaborating among themselves to improve analysis and interpretation of NAEP ELL data and state achievement data given NCLB goals.[5]

[5] In 2005, Eric Zilbert, the NAEP coordinator for California, and Pete Bylsma, the Director for Research and Evaluation for Washington State, participated in a Council of Chief State School Officers Annual Assessment meeting session on the topic of challenges and opportunities in using ELL NAEP scores to improve interpretation of state assessment scores under NCLB.

References

Abedi, J. (2004a). Inclusion of students with limited English proficiency in NAEP: Classification and measurement issues. CSE Report 629, CRESST/University of California, Los Angeles.

Abedi, J. (2004b). The No Child Left Behind Act and English language learners: Assessment and accountability issues. *Educational Researcher*, *33* (1), 4-14.

Fuhrman, S. and Elmore, R. (Eds.) (2004). Redesigning accountability systems for education. New York: Teachers College Press.

Haertel. E. (December 2003). Including students with disabilities and English language learners in NAEP: Effects of differential inclusion rates on accuracy and interpretability of findings. Commission Paper for the *NAGB Conference on Increasing the Participation of SD and LEP Students in NAEP*.

Hakuta, K., Butler, Y., and Witt, D. (2000). How Long Does It Take English Learners to Attain Proficiency? Policy Report 2000- 1, Santa Barbara: University of California, Santa Barbara, The University of California Linguistic Minority Research Institute.

Jones, L and Olkin, I (2004). The Nation's report card: Evolution and perspectives. Bloomington, Indiana: Phi Delta Kappa Educational Foundation.

Koenig, J. and Bachman, L. (2004). Keeping score for all. The effects of inclusion and accommodation policies on large-scale educational assessments. National Research Council. Board on Testing and Assessment. Committee on Participation of English Language Learners and Students with Disabilities in NAEP and Other Large-Scale Assessments. Washington DC: The National Academies Press.

Marion, S., White, C., Carlson, D., Erpenbach, W., Rabinowitz, S., and Sheinker, J. (December 2002). Making valid and reliable decisions in determining adequate yearly progress. Washington DC: Council of Chief State School Officers.

Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational Measurement* (3rd ed.) New York: Macmillan, pp. 13-13.

National Center for Education Statistics (last updated 3 July 2006). NAEP Inclusion Policy. Inclusion of Special-Needs Students (retrieved 10-23-06 from http://nces.ed.gov/nationsreportcard/about/inclusion.asp).

National Clearinghouse for English Language Acquisition and Language Instruction Educational Programs. NCELA FAQ: How has the English language learner (ELL) population changed in recent years? (retrieved 2-26-06 from http://www.ncela.gwu.edu/expert/faq/08leps.html).

Perie, M., Grigg, W., and Dion, G. (2005). The nation's report card: Mathematics 2005 (NCES 2006–453). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Perie, M., Grigg, W., and Donahue, P. (2005). The nation's report card: Reading 2005 (NCES 2006–451). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

U.S. Department of Education (October 2006). LEP partnership. (Retrieved 11-16-06 from http://www.ed.gov/about/inits/ed/lep-partnership/index.html)

U.S. Department of Education (July 2006). Building partnerships to help English language learners. (Retrieved 11-16-06 from http://www.ed.gov/nclb/methods/english/lepfactsheet.html)

U.S. Department of Education (2005). Biennial Evaluation Report to Congress on the Implementation of the State Formula Grant Program. 2002-2004 English Language Acquisition, Language Enhancement and Academic Achievement Act (ESEA, Title III, Part A). Washington DC.

U.S. Department of Education (October 29, 2004) Letter to Chief State School Officers regarding the NCLB Standards and Assessment Peer Review Process (retrieved 12-28-04 from http://www.ed.gov/)

U.S. Department of Education (February 2004). Secretary Paige Announces New Policies to Help English Language Learners. (Retrieved 2-26-06 from http://www.ed.gov/news/pressreleases/2004/02/02192004.html).

U.S. Department of Education (last updated February 19, 2004). Fact Sheet: NCLB Provisions Ensure Flexibility and Accountability for Limited English Proficient Students (retrieved 12-12-04 from http://www.ed.gov/nclb/accountability/schools/factsheet-english.html).

Zilbert, E. (November 19. 2004). Comparing Performance on NAEP with Growth on the SAT 9 and California Standards Test in English Language Arts. Paper presented at the California Educational Research Association Conference. Pasadena California.