

NCLB Assessment Issues for English Language Learners

Teachers of English to Speakers of Other Languages

Jamal Abedi

National Center for Research on Evaluation,
Standards, and Student Testing

*UCLA Graduate School of Education & Information Studies
April 2, 2005*



NCLB Accountability System

States should implement high-quality annual assessments in reading/language arts, mathematics, and later in science that:

1. Apply the same high standards of academic achievement to all public schools
2. Are statistically valid and reliable
3. Result in substantial academic improvement for all students



What is AYP?

The predetermined increase in the percentage of students scoring at the "proficient" level or higher on Annual Measureable Objectives (AMOs) is referred to as Adequate Yearly Progress (AYP).

Each state establishes a timeline for all students to reach "proficient" or higher, no more than 12 years after the 2001–2002 start date.

The first increase occurs within the first two years.



How Should AYP be Reported?

AYP is reported for schools, districts, and states for all students, as well as for the following student subgroups:

1. Economically disadvantaged
2. Major racial and ethnic groups
3. Students with disabilities
4. Students with limited English proficiency



Technical Issues in AYP Reporting

- **Defining academic achievement**
- **Alignment of content standards with test items**
- **Setting achievement levels**



Corrective Action

Year 1

- Revise school plan
- Use 10% of funds for staff development
- Provide school choice with paid transportation
- District provides technical assistance (TA)

(District J Special Administrators Academy
Session, August 13, 2003)

Corrective Action

<p>Year 2</p> <ul style="list-style-type: none"> ◆ <i>Continue</i> <ul style="list-style-type: none"> • Staff development • School choice • District TA ◆ <i>Add</i> <ul style="list-style-type: none"> • Supplemental services/tutoring 	<p>Year 3</p> <ul style="list-style-type: none"> ◆ <i>Continue</i> <ul style="list-style-type: none"> • District TA • School choice • Supplemental Services ◆ <i>Add</i> <ul style="list-style-type: none"> • District corrective action
---	---

Corrective Action

<p>Year 4</p> <ul style="list-style-type: none"> ◆ <i>Continue</i> <ul style="list-style-type: none"> • District TA • School choice • Supplemental Services ◆ <i>Add</i> <ul style="list-style-type: none"> • Develop alternative governance plan 	<p>Year 5</p> <ul style="list-style-type: none"> ◆ <i>Implement alternative governance plan</i> <ul style="list-style-type: none"> • Reopen as charter • Replace staff • Contact with external entity • Takeover by state
--	--

Problems in AYP Reporting: Focus on LEP Students

1. Problems in classification/reclassification of LEP students (moving target subgroup)
2. Measurement quality
3. Low baseline
4. Instability of the LEP subgroup
5. Sparse LEP population
6. LEP cutoff points (Conjunctive vs. Compensatory model)

Issues with LEP Assessment, Accommodation, and AYP Reporting

1. Problems in Classification/ Reclassification of LEP Students

- ◆ LEP classification criteria varies
- ◆ Discrepancies in LEP criteria cause inconsistencies nationwide
- ◆ Accuracy of AYP reporting for LEP students suffers

Issues with LEP Assessment, Accommodation, and AYP Reporting

2. Measurement quality of AYP instruments for LEP students

- ◆ Students' yearly progress is measured by performance on academic achievement tests
- ◆ Academic achievement tests are constructed and normed for native speakers
- ◆ These tests have lower reliability and validity for LEP populations
- ◆ Therefore, test results should not be interpreted for LEP students as for non-LEP students

Site 2 Stanford 9 Sub-scale Reliabilities (1998) Grade 9 Alphas

Sub-scale (Items)	Non-LEP Students			FEP	RFEP	LEP
	Hi SES	Low SES	English Only			
Reading, N=	205,092	35,855	181,202	37,876	21,869	52,720
-Vocabulary (30)	.828	.781	.835	.814	.759	.666
-Reading Comp (54)	.912	.893	.916	.903	.877	.833
Average Reliability	.870	.837	.876	.859	.818	.750
Math, N=	207,155	36,588	183,262	38,329	22,152	54,815
-Total (48)	.899	.853	.898	.898	.876	.802
Language, N=	204,571	35,866	180,743	37,862	21,852	52,863
-Mechanics (24)	.801	.759	.803	.802	.755	.686
-Expression (24)	.818	.779	.812	.804	.757	.680
Average Reliability	.810	.769	.813	.803	.756	.683
Science, N=	163,960	28,377	144,821	29,946	17,570	40,255
-Total (40)	.800	.723	.805	.778	.716	.597
Social Science, N=	204,965	36,132	181,078	38,052	21,967	53,925
-Total (40)	.803	.702	.805	.784	.722	.530

Grade 11 Stanford 9 Reading and Science Structural Modeling Results (DF=24), Site 3

	All Cases (N=7,176)	Even Cases (N=3,588)	Odd Cases (N=3,588)	Non-LEP (N=6,932)	LEP (N=244)
Goodness of Fit					
Chi Square	1786	943	870	1675	81
NFI	.931	.926	.934	.932	.877
NNFI	.898	.891	.904	.900	.862
CFI	.932	.928	.936	.933	.908
Factor Loadings					
Reading Variables					
Composite 1	.733	.720	.745	.723	.761
Composite 2	.735	.730	.741	.727	.713
Composite 3	.784	.779	.789	.778	.782
Composite 4	.817	.722	.712	.716	.730
Composite 5	.633	.622	.644	.636	.435
Math Variables					
Composite 1	.712	.719	.705	.709	.660
Composite 2	.695	.696	.695	.701	.581
Composite 3	.641	.628	.654	.644	.492
Composite 4	.450	.428	.470	.455	.257
Factor Correlation					
Reading vs. Math	.796	.796	.795	.797	.791

Note. NFI = Normed Fit Index. NNFI = Non-Normed Fit Index. CFI = Comparative Fit Index.

Issues with LEP Assessment, Accommodation, and AYP Reporting

3. Low LEP Baseline Scores

- Schools with high numbers of LEP students have lower baseline scores
- For those schools, year-to-year progress goals are much more challenging and might be considered unrealistic
- Their students may continue to struggle with the same academic disadvantages and limited school resources

Issues with LEP Assessment, Accommodation, and AYP Reporting

4. Instability of LEP Subgroup

- A student's LEP status is not stable over time, it is a moving target
- When a student improves to "proficient," the student leaves the LEP subgroup
- Those who remain are low performing
- New arrivals with even lower levels of language proficiency may join the subgroup
- Even with the best resources, the LEP subgroup AYP indicator has little chance of rising

Issues with LEP Assessment, Accommodation, and AYP Reporting

5. Sparse LEP Population

- The number of LEP students varies across the nation
- A minimum of 25 is needed for reliable AYP reporting (Linn, Baker, & Herman, 2002)
- To detect moderate changes, several hundred subjects may be needed (Hill & DePascale, 2003)
- In many states and districts, the number of LEP students is not enough for any meaningful analyses
- This might skew some states' accountability and adversely affect state and federal policy decisions

Issues with LEP Assessment, Accommodation, and AYP Reporting

6. LEP Cutoff Points (Conjunctive vs. Compensatory Model)

- Earlier legislation (e.g., Improving America's Schools Act, IASA), adopted a Compensatory model
- In a Compensatory model, higher scores in content areas with less language demand compensate for scores in areas with higher language demand
- NCLB AYP reporting is based on a conjunctive model in which students should score at a "proficient" level in all required content areas
- This makes all the AYP requirements more difficult for schools with many LEP students

Normal Curve Equivalent Means & Standard Deviations for Students in Grades 10 and 11, Site 3 School District

	Reading		Science		Math	
	M	SD	M	SD	M	SD
Grade 10						
SWD only	16.4	12.7	25.5	13.3	22.5	11.7
LEP only	24.0	16.4	32.9	15.3	36.8	16.0
LEP & SWD	16.3	11.2	24.8	9.3	23.6	9.8
Non-LEP/SWD	38.0	16.0	42.6	17.2	39.6	16.9
All students	36.0	16.9	41.3	17.5	38.5	17.0
Grade 11						
SWD Only	14.9	13.2	21.5	12.3	24.3	13.2
LEP Only	22.5	16.1	28.4	14.4	45.5	18.2
LEP & SWD	15.5	12.7	26.1	20.1	25.1	13.0
Non-LEP/SWD	38.4	18.3	39.6	18.8	45.2	21.1
All Students	36.2	19.0	38.2	18.9	44.0	21.2

Note. LEP = limited English proficient. SWD = students with disabilities.



1. Problems in Classification/ Reclassification of LEP Students

Recommendations

- Use multiple criteria that are valid and highly relevant to LEP classification
- Develop English proficiency measures based on English proficiency and academic content standards
- Test the newly developed measures with sufficient field data to demonstrate validity
- Use valid cut-scores through multiple standard-setting approaches based on both examinees and instruments



2. Measurement Quality of AYP Instruments for LEP Students

Recommendations

- Identify nuisance or confounding variables (e.g., linguistic complexity of test items) that may negatively affect measurement outcome of content-based tests
- In high-stakes assessment, specify construct being measured and separate it from other (construct-irrelevant) factors
- Identify & provide valid accommodations (e.g., language-related) in both instruction & assessment of English learners



3. Low LEP Baseline Scores

Recommendations

Provide more resources to schools with lower baseline scores, for example:

- Hire more qualified teachers
- Provide professional training in areas of need
- Provide after-school programs, student tutoring, and activities shown to help low-performing students
- Provide diagnostic information for both teachers and parents



4. Instability of LEP Subgroup

Recommendations

- When LEP students achieve an English proficiency level qualifying them for reclassification, keep them in the LEP subgroup only for accountability purposes
- Instead of current exceptions to LEP AYP reporting (keeping LEPs in main cohort for some time), we propose a semi-cohort approach to LEP subgroups
- Provide more resources (professional training, aides, smaller class size) for teachers of new LEP students with substantially lower English proficiency levels than those already in the semi cohort



5. Sparse LEP Population

Recommendations

- Small numbers of LEP students should not affect accountability requirements
- While any AYP based on a small number of LEP students may not be reliable, schools, districts, and states must still be accountable
- States should provide help to schools with low-performing LEP students, regardless of the school's LEP population



6. LEP Cutoff Points (Conjunctive vs. Compensatory Model)

Recommendations

- A compromise approach should be considered
- Research is needed to set a series of cut-scores that provide a more valid and flexible decision-making process



More Information

These issues have been discussed in more details in the following:

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

Any Questions:

Contact Jamal Abedi at:

jabedi@cse.ucla.edu or (310) 206-4346