Issues in the Development of Annual Measurable Achievement Objectives (AMAOs) for WIDA Consortium States

H. Gary Cook
Value-Added Research Center, Wisconsin Center for Education Research,

Tim Boals
Carsten Wilmes
Martin Santos
WIDA Consortium, Wisconsin Center for Education Research

We are indebted to WIDA Consortium states for their support and feedback on this document. We would like to especially thank Margo Gottlieb and Robert Linquanti for their helpful comments. Any mistakes, errors or omissions are those of the authors.

Table of Contents

List of Tables

List of Figures

Purpose of Report and Background

Title III of the No Child Left Behind (NCLB) Act (2001) requires State Education Agencies (SEAs) to develop progress and attainment benchmarks, called Annual Measurable Achievement Objectives (AMAOs) for English language learners (ELLs). AMAOs must be based on annual assessments of English proficiency in the domains of listening, speaking, reading, writing, and comprehension (derived from listening and reading). In WIDA states, the NCLB approved English language proficiency assessment is *Assessing Comprehension and Communication in English State to State*, ACCESS for ELLs®. This assessment is aligned to WIDA English proficiency standards and reflective of the state specific academic content standards of member states as required under the Act.

As a consortium dedicated to assisting member states with both legal compliance and improving the education of ELLs, WIDA seeks to provide guidance and support to states and local schools. The goal of this paper is to assist states as they formulate AMAO targets, specifically this report focuses on AMAO 1 and 2 targets. WIDA issued two papers prior to this on the same topic. Margo Gottlieb was the primary author of the first paper entitled *Using ACCESS for ELLs® Data at the State Level: Considerations in Reconfiguring Cohorts, Resetting Annual Measurable Achievement Objectives (AMAOs), and Redefining Exit Criteria for Language Support Programs Serving English Language Learners (DRAFT).* This work became WIDA Technical Report #3, and was made available to SEAs in January 2006. At that time, the Consortium had only one year's

English proficiency data based on three states (AL, ME, VT) that tested in spring 2005. The recommendations were tentative and based largely on national studies estimating average time ELLs typically need to reach full English proficiency given the linguistic demands of academic subject areas. The following year, WIDA asked Gary Cook, researcher at the Wisconsin Center for Education Research, to analyze ELL growth profiles based on two years of available data in the same three states. This report was issued August 30, 2006 as *Composite Score Growth Profile Based on Years 1 & 2 ACCESS for ELLs® Data, WIDA Consortium Draft Recommendations.* That analysis represented a significant refinement of the general recommendations in the first report.

Now, WIDA has three years of results from the original three states and two years of results from nine additional states that tested in 2006 and 2007. These data give us the opportunity to further refine our progress estimates and discuss the more complex growth patterns that are emerging. WIDA also issued a Bridge Study Report by Dorry Kenyon in August 2006 (Technical Report #2), available in the technical reports section of the ACCESS for ELLs® webpage on the Consortium website (www.wida.us). The Bridge Study Report compares four prior ELP assessments (IPT, LAS, MAC II, and Woodcock-Muñoz) to scores on ACCESS for ELLs®. The Bridge data were used by at least three WIDA states when setting their initial AMAOs for ACCESS for ELLs®.

WIDA Standards and Assessments

WIDA's English Language Proficiency Standards, and corresponding Model Performance Indicators, were written to address both NCLB Title III requirements and

the growing awareness that support for learning English within K-12 school settings must focus more specifically on the language demands of academic content area classrooms. By the early 1980s, the construct we now refer to as "academic language proficiency" emerged, albeit with less empirical support than today (Cummins, 1983). In 1979, Mohan's article "Language Teaching and Content Teaching" was published in TESOL Journal. Mohan's seminal text *Language and Content* was published in 1986, and the modern movement to teach language through content subject matter had begun.

The heightened awareness of the differing demands of academically oriented English coincided with the recognition that the process for gaining full English language proficiency was likely much longer. Students who seemed to carry on good conversations in English within a year or two struggled using English effectively in academic classes for a substantial time longer (up to 7 years). Some students with weaker literacy skills in their primary language or limited school experience appeared to require even more time to close the academic content and language gaps. This recognition fueled the need for teaching language through content since English language learners could not afford large chucks of the day dedicated to mastering "English" while other students in the school had more time on task for mathematics, science, and other subjects. By the time NCLB came in 2001, the force of federal law backed the idea that ELLs must be included in school level accountability, and hence assessments of both academic content and English proficiency. Gone were the days when you could afford to teach English and worry about catching up with everything else later.

Prior to NCLB, practical tools to assist teachers in making the transition to teaching grade appropriate content along with subject specific English were scarce. NCLB's requirement that states develop English proficiency standards aligned to academic language and linked to academic content standards has served to provide "blueprints" for materials development, lesson planning, classroom assessment and staff development. The English proficiency standards also serve as the blueprints for English proficiency tests. Ideally, all standards and assessments within the system push schools in the same direction, i.e., including ELLs in grade appropriate curriculum taught in ways that maximizes their learning of content while increasing proficiency in the English. This is what WIDA's English language proficiency standards seek to accomplish.

To assess WIDA standards, the Consortium has developed an assessment called ACCESS for ELLs® (ACCESS). ACCESS is composed of four domain based tests (speaking, listening, reading, and writing) assessing students at a variety of grade-level clusters: Kindergarten, 1st and 2nd grade, 3rd through 5th grade, 6th through 8th grade, and 9th through 12th grade levels. For each domain, grade level cluster (except K), ACCESS is a three tier based assessments: A, B, and C. Tier A assessments are designed to target children at the three lowest language proficiency levels (Levels 1-3). Tier B assessments are directed toward students at the middle proficiency levels (Levels 2-5), and Tier C assessments measure students at the higher end of the proficiency spectrum (Levels 3-6). Psychometrically, ACCESS is vertically scaled across grades and proficiency levels using RASCH, IRT scaling methodology. The ACCESS scale ranges from 100 and 600. Because of the vertical scaling design, it is possible to longitudinally monitor students'

progress in English across grades and proficiency levels. ACCESS is a highly reliable assessment designed specifically to assess the academic language proficiency as specified in the WIDA standards. For more information on the reliability and validity of ACCESS see http://www.wida.us/assessment/ACCESS_techReports/index.aspx.

Data Used for Analyses

The data used for this report come from 3 WIDA Consortium states over a three year period (2005, 2006 and 2007) across all grades (Kindergarten to 12th grade). A total of 12,836 ELL students are included in analyses, 9,542 from State A, 2,154 from State B, and 1,140 from State C. In 2005, all 12,836 students are included in the sample. A total of 12,014 students are retained in the 2006 sample, and 9,353 students remain in the 2007 sample. Reduction in student numbers over school years result from students exiting ELL programs, moving, missing the assessment, or missing data. It is important to note that this is a longitudinal dataset, i.e., students included in this sample have at least 2 data points, and over 9,000 have three data points. These trend data provide the opportunity to examine how students grow in language proficiency within their cohorts.

The ACCESS assessment provides a variety of domain and composite scores to aid in interpreting students' academic language proficiency. The following table outlines the scores provided by the ACCESS assessment.

Table 1: ACCESS for ELLs Weighted Scores

Overall Composite	Contribution of Language Domains by Percent			
	Listening	Speaking	Reading	Writing
Oral Language	50%	50%	–	–
Literacy	–	–	50%	50%

Comprehension	30%	–	70%	–
Overall	15%	15%	35%	35%

Table taken from the ACCESS, Interpretive Guide for Score Reports (Table 1).
http://www.wida.us/assessment/ACCESS%20Interpretive%20Guide07.pdf

RASCH model generated scale scores are created for the four domain tests. Domain test scales score are then reported to students, parents, teachers, schools, and districts in a variety of ways (proficiency levels, proficiency level decimals, and scale scores). The most general score provided by ACCESS is the Overall Composite. This Overall Composite (here onward termed Composite) is a domain, weighted score, interpreted from scale scores, derived in the following manner: Speaking (15%), Listening (15%), Reading (35%), and Writing (35%). Thus 70% of the Composite score associates with text-based proficiency and 30% associates with oral/aural proficiency. All AMAO analyses reported in this paper will use the Composite score exclusively since most WIDA states use this as the metric for AMAO expectations (see Appendix A).

Development of Annual Measurable Achievement Objective 1: Progress

A premise behind AMAO 1, the number or percent of English language learners making marked progress targets, is the continuous improvement of districts charged with teaching English to English language learners (ELLs). Two elements associate with this improvement: first, that students make appropriate progress in their language learning, and second, that district ELL programs continue to improve the rate at which students make progress. This section of the report outlines a process for establishing AMAO 1 goals using data from three WIDA states, as described earlier. All three states' ACCESS Composite scores are combined to form one dataset. The methods for establishing AMAO 1 criteria outlined below are designed to act as guides in assisting Consortium members in setting their own AMAO 1 criteria. Prior to discussing this process, it would be helpful to review what we know about ELL students' growth in English.

What we know from second language acquisition research

What do we know about the second language acquisition of students for whom AMAO policies are created? ***In our view, setting AMAO policies without understanding how second language learning occurs is fraught with problems***. For example, a state may set student growth expectations that are well beyond what is acquisitionally possible. We may want students to grow faster in their acquisition of English, but as seen in second language acquisition research, some linguistic features take time to master. Demanding that students or language programs "move students along," in some cases, belies the very nature of language learning and are unreasonable and unrealistic. The converse is also

true, student growth expectations could be set abysmally low, so low in fact that students would take inordinately long to learn English. According to Scarcella (2003), such low expectations have severe economic consequence and do a disservice to students. The goal is to find expectations that are reasonable but challenging. We believe this can be achieved by keeping in mind what we know about child second language acquisition and by examining real students' data, as expressed on ACCESS. Both approaches provide information to set reasonable, realistic expectations.

It is beyond the scope of this paper to overview child second language research in detail. Research suggests that students' second language acquisition is influenced by a variety of factors, e.g., social context, language aptitude, types of inputs and interactions, cognitive processing, influence of first language, prior educational experiences. For those interested in learning more about child and adult second language acquisition see McLaughlin (1984) and Doughty & Long (2003). Two particular factors are of specific interest in setting AMAOs:

- The maturational constraints to learning a second language (see Hyltenstam & Abrahamsson, 2003, Collier, 1988, 1995), and
- The stabilization of language development (Long, 2003).

To the first point, second language acquisition (SLA) researchers have observed that "children are more efficient language learners than adults…" (Hyltenstam & Abrahamsson, 2003, p.539). A variety of theories have been proposed to explain why that proposition is so. One popular hypothesis is that there is a "critical period" in which language learning can occur with little effort, typically before puberty. However, the

critical theory hypothesis is by no means a consensus opinion. Many argue against this idea and suggest that adults can acquire language fluently, but at dramatically different rates. Others have suggested that the learning environments themselves play a larger role in facilitating acquisition, i.e., a Kindergarten class provides more language interaction than a high school lecture. (Krashen, 1987) Another important caveat is that stronger literacy skills in a student's native language, more likely seen in older children but not always, will make a significant difference in rate of second language acquisition (Collier, 1995). These caveats aside, most agree that younger learners learn at higher rates than older learners.

To the second point, researchers observe that as language learners move to higher levels of proficiency, the rates at which language is acquired slows down. In many instances, language learning can fossilize, or a Long (2003) suggests stabilize. That is, language learners operate at an interlanguage (Selinker, 1972) level. Here interlanguage refers to something between the first and second language. It is not the fluent use of the second language, nor is it only first language structures; it is something in between. Moving through interlanguage stages (or the second language acquisition continuum) can be protracted. The mastery of complex language features often requires a period of "germination" or trial and error. There is debate whether children "stabilize" in their language development or at what age stabilization (or fossilization) might occur. Regardless, the observational reality is that second language learners at higher levels of English language proficiency require more time to master linguistic features than lower level language learners.

These two factors should be considered when setting AMAO progress expectations. Both factors can be combined into the following conceptualization: *lower is faster, higher is slower*. That is, language learners in lower grades (younger students) acquire language at faster rates. Students at lower proficiency levels acquire language at faster rates. Taken together if two students were at the same proficiency level but at different grades, we would predict that the younger student would grow at a higher rate than the older student. This conceptualization should color AMAO expectations.

A model for setting Annual Measurable Achievement Objectives

The model we recommend and will be using for setting AMAOs is that set forth by Linquanti & George (in press). In this paper, they describe the process that California Department of Education staff and outside consultants engaged in to establish California's AMAO criteria using data from the CELDT (California English Language Development Test). The federal statutes that Linquanti and George base their procedures on are as follows:

AMAO Definitions

Section 3122(a)(1): *Each State educational agency or specially qualified agency receiving a grant under subpart 1 shall develop annual measurable achievement objectives for limited English proficient children served under this part that relate to such children's development and attainment of English proficiency while meeting*

challenging State academic content and student academic achievement standards as required by section 1111(b)(1).

AMAO Categories

AMAO 1 – Progress (3122(a)(3)(A)(i): *at a minimum, annual increases in the number or percentage of children making progress in learning English*;

AMAO 2 – Proficiency (3122(a)(3)(A)(ii): *at a minimum, annual increases in the number or percentage of children attaining English proficiency by the end of each school year, as determined by a valid and reliable assessment of English proficiency consistent with section 1111(b)(7); and*

AMAO 3 – AYP (3122(a)(3)(A)(iii): *making adequate yearly progress for limited English proficient children as described in section 1111(b)(2)(B)*;

Linquanti and George set forth five key decisions needed to establish AMAO 1 and AMAO 2 expectations. The five AMAO 1 decisions are as follows:

1. Determine scoring metric to be used to measure growth,
2. Determine annual growth target,
3. Set the starting point for AMAO 1 targets,
4. Set the ending point for AMAO 1 targets,
5. Determine annual rate of growth.

(Linquanti & George, in press, pp.2-3)

The headings in this section are based upon the five key decisions above.

Determine Scoring Metric

The ACCESS assessment provides three potential score metrics for AMAO 1:

- WIDA English Language Proficiency Levels 1-5 (Entering, Beginning, Developing, Expanding, Bridging)
- WIDA English Language Proficiency Level Decimal Scores (1.1 to 5.9), where

  The whole number indicates the student's language proficiency level as based on the WIDA English Language Proficiency Standards. The decimal indicates the proportion within the proficiency level range that the student's scale score represents, rounded to the nearest tenth. Proficiency level scores do not represent interval data. The interval between corresponding scale scores for 2.2 to 3.2, for example, are not necessarily the same as between a 3.2 and a 4.2.[1]
- Scale Scores (100-600)-Rasch, IRT vertically scaled scores from Kindergarten to 12th grade ranging from 100 to 600.

Both proficiency levels and proficiency level decimals were set using traditional standard setting techniques. For more information on this process see the WIDA Technical Manuals at http://www.wida.us/assessment/ACCESS_techReports/index.aspx. The WIDA proficiency scores used in this analysis are the most recent, grade specific

[1] WIDA Consortium. (2007). *ACCESS for ELLs® Interpretive Guide for Score Reports, Spring 2007*. The Board of Regents of the University of Wisconsin System; Madison, WI.

cutpoints that were reset in 2007. For analyses new WIDA proficiency cut scores (based on individual grades rather than grade level clusters) are back-applied to 2005 and 2006 data. This analysis provides apples-to-apples comparisons.

Analyses reported here use only ACCESS Composite scores. Again, the Composite score is weighted version of the four ACCESS domain scores (Speaking 15%, Listening 15%, Reading 35%, and Writing 35%).

Since WIDA Consortium members' AMAO policies vary in the metrics used to determine AMAO 1 (and AMAO 2), all three metrics are analyzed. See the Appendix A for samples of state policies. For clarity in presentation, all three metrics will be presented within each heading.

Determine Annual Growth Target

Proficiency Level Gain

The following series of tables, graphs and figures show within student (i.e. within cohort) growth across the 2005, 2006 and 2007 school years aggregated across analyzed states. States that used the proficiency level metric often set AMAO 1 expectations based upon the percent of students who gain one or more proficiency levels per year. The following tables show the percent of students who gain one or more WIDA proficiency level by grade band and proficiency level.

Table 2: Percent of Students Gaining One or More Proficiency Levels--2005 to 2006

Grade Band	Initial Composite Score Proficiency Level				
	1	2	3	4	5

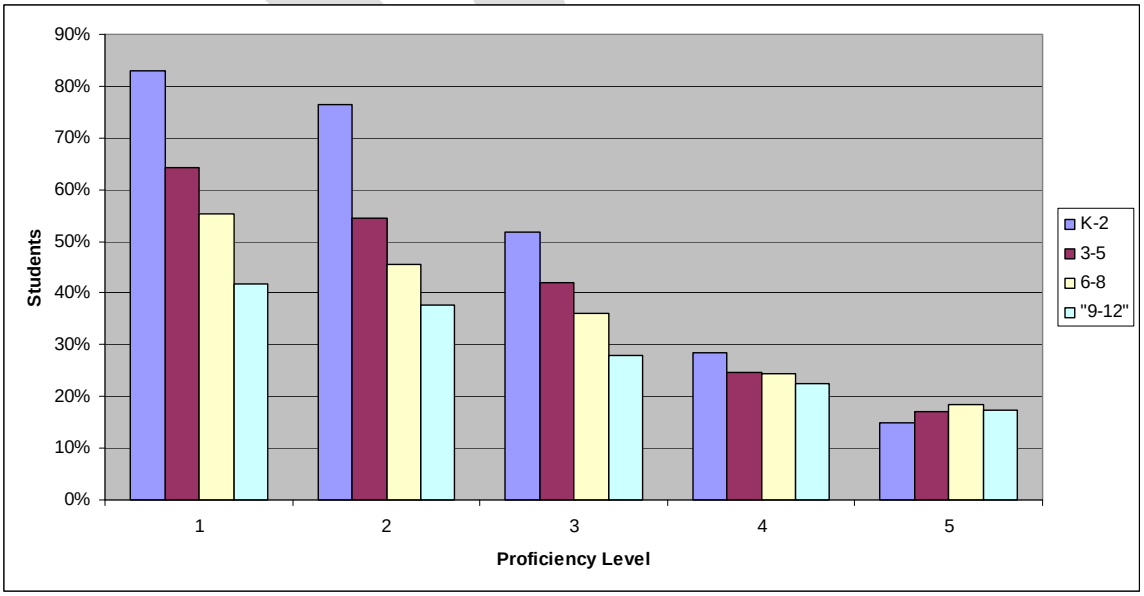
K-2	84%	79%	52%	29%	13%
3-5	60%	53%	40%	26%	19%
6-8	55%	48%	38%	27%	26%
9-12	39%	39%	29%	25%	21%

Table 3: Percent of Students Gaining One or More Proficiency Levels--2006 to 2007

Grade Band	Initial Composite Score Proficiency Level				
	1	2	3	4	5
K-2	79%	70%	52%	28%	20%
3-5	74%	57%	44%	23%	14%
6-8	57%	42%	34%	22%	8%
9-12	46%	36%	26%	20%	13%

Tables 2 and 3 highlight expectations based on child second language theory, i.e., lower is faster, higher is slower. Lower grades and proficiency levels have higher percentages of students gaining one or more proficiency levels per year. Figure 1 graphically displays this characteristic by aggregating both years together.

Figure 1: Aggregate Percent of Students Gaining 1 or More Proficiency Levels 2005-2007

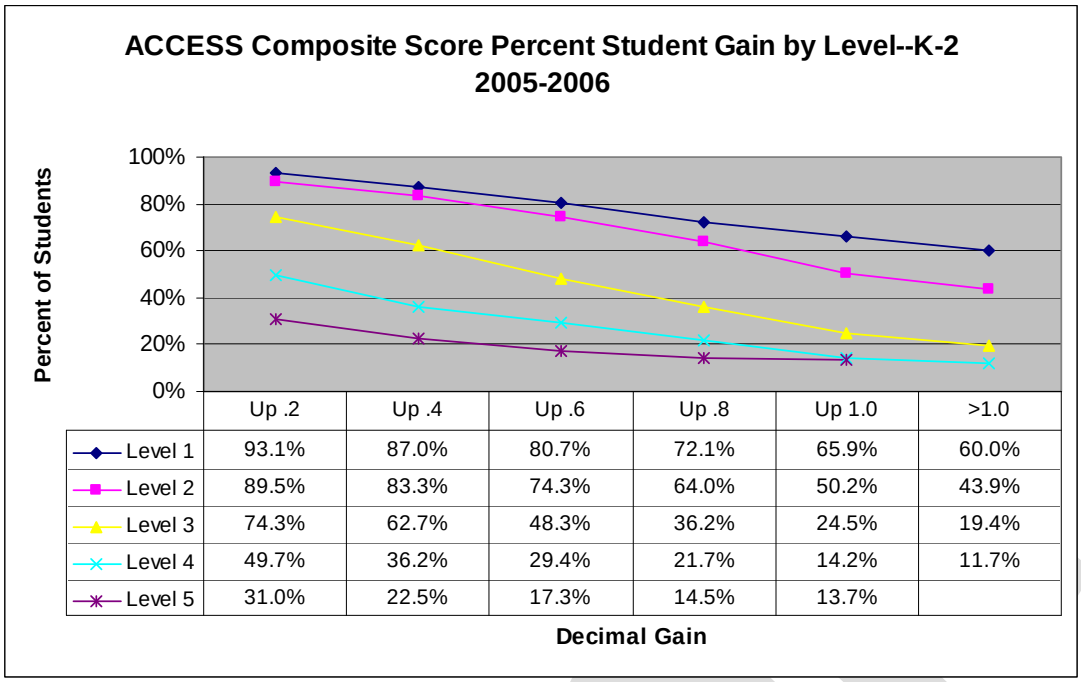


If the same AMAO 1 expectations were set across proficiency levels and grade bands, students at higher grades or higher proficiency levels would not gain at the same rate as their lower grade or proficiency level classmates. This suggests the need for setting different targets at different grade bands and proficiency levels.

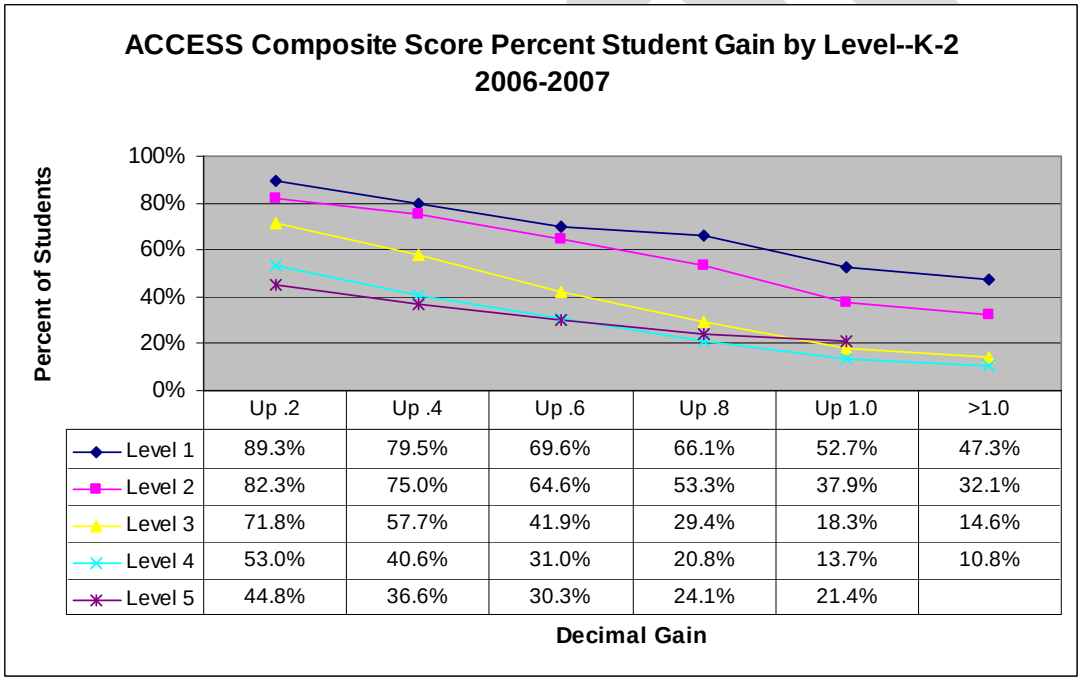
Proficiency Level Decimal Gain

Some states set AMAO targets based on proficiency level decimal increases. The following series of figures show the percent of students gaining in composite proficiency levels in 0.2 decimal increments.

Figure 2: Composite Score Proficiency Gain in 0.2 Increments-K-2 Grade Band

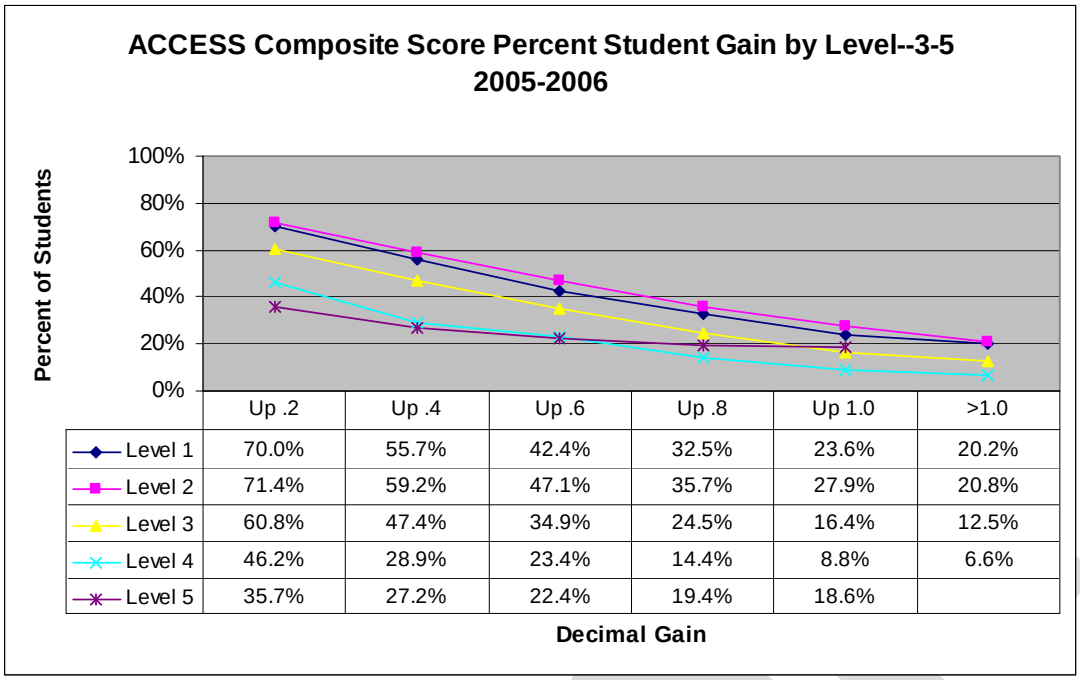


	Up .2	Up .4	Up .6	Up .8	Up 1.0	>1.0
Level 1	93.1%	87.0%	80.7%	72.1%	65.9%	60.0%
Level 2	89.5%	83.3%	74.3%	64.0%	50.2%	43.9%
Level 3	74.3%	62.7%	48.3%	36.2%	24.5%	19.4%
Level 4	49.7%	36.2%	29.4%	21.7%	14.2%	11.7%
Level 5	31.0%	22.5%	17.3%	14.5%	13.7%	

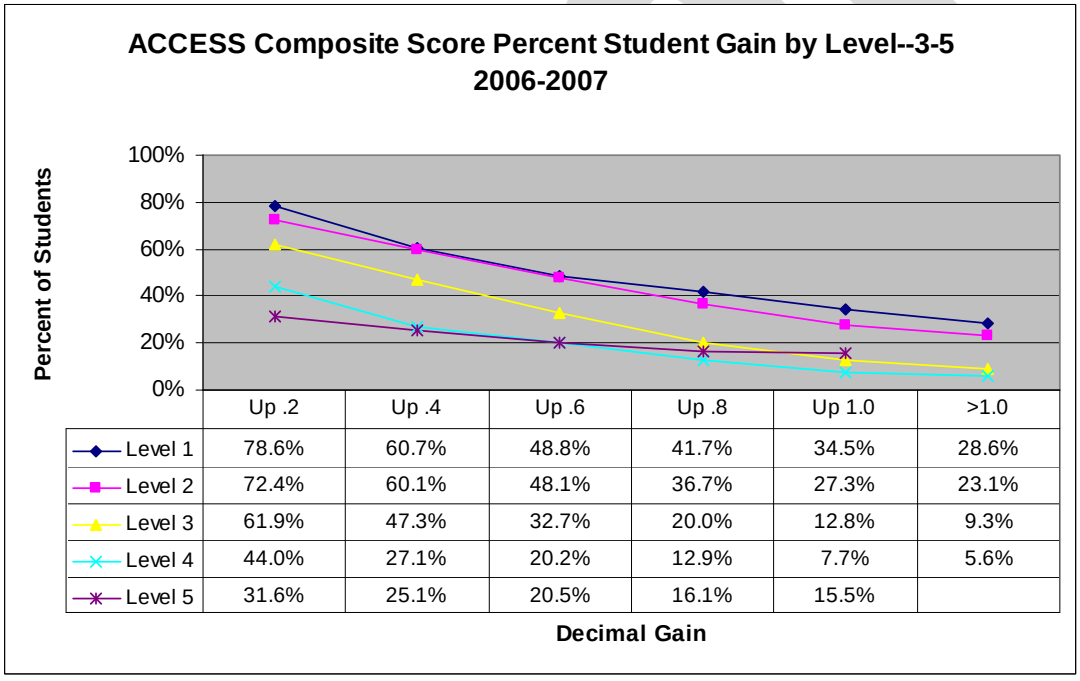


	Up .2	Up .4	Up .6	Up .8	Up 1.0	>1.0
Level 1	89.3%	79.5%	69.6%	66.1%	52.7%	47.3%
Level 2	82.3%	75.0%	64.6%	53.3%	37.9%	32.1%
Level 3	71.8%	57.7%	41.9%	29.4%	18.3%	14.6%
Level 4	53.0%	40.6%	31.0%	20.8%	13.7%	10.8%
Level 5	44.8%	36.6%	30.3%	24.1%	21.4%	

Figure 3: Composite Score Proficiency Gain in 0.2 Increments-3-5 Grade Band

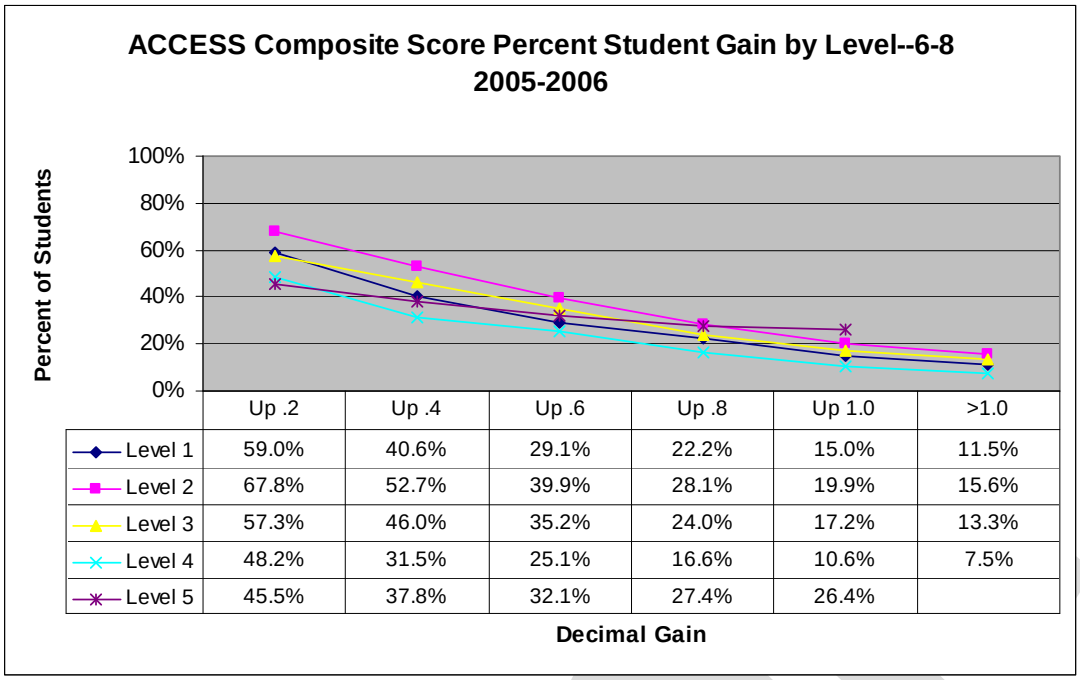


	Up .2	Up .4	Up .6	Up .8	Up 1.0	>1.0
Level 1	70.0%	55.7%	42.4%	32.5%	23.6%	20.2%
Level 2	71.4%	59.2%	47.1%	35.7%	27.9%	20.8%
Level 3	60.8%	47.4%	34.9%	24.5%	16.4%	12.5%
Level 4	46.2%	28.9%	23.4%	14.4%	8.8%	6.6%
Level 5	35.7%	27.2%	22.4%	19.4%	18.6%	

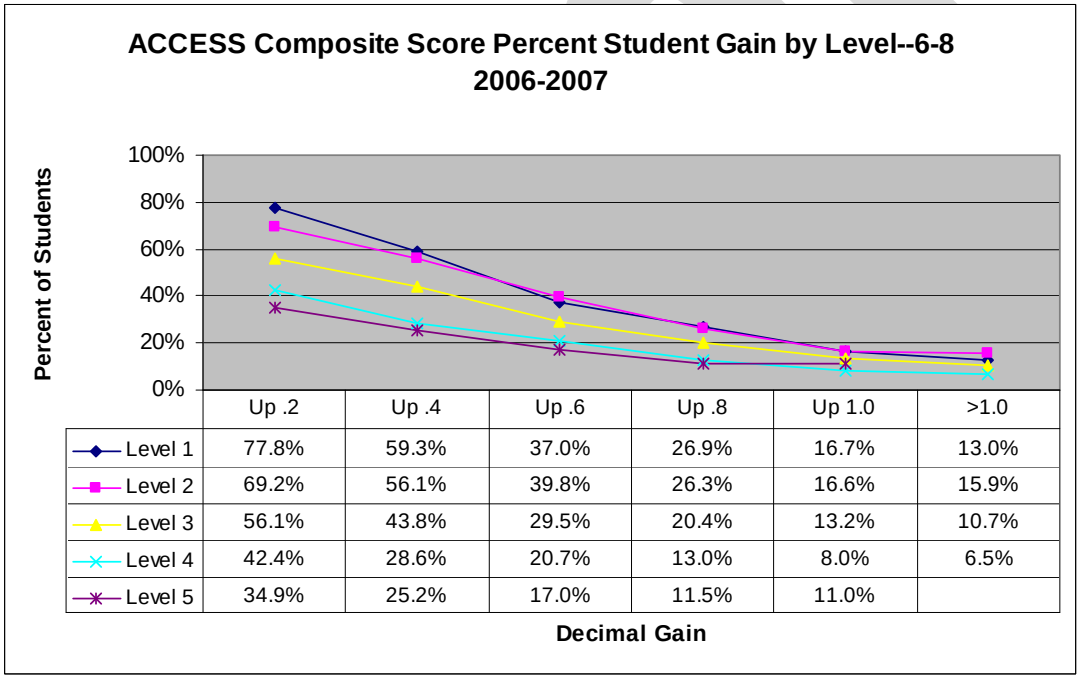


	Up .2	Up .4	Up .6	Up .8	Up 1.0	>1.0
Level 1	78.6%	60.7%	48.8%	41.7%	34.5%	28.6%
Level 2	72.4%	60.1%	48.1%	36.7%	27.3%	23.1%
Level 3	61.9%	47.3%	32.7%	20.0%	12.8%	9.3%
Level 4	44.0%	27.1%	20.2%	12.9%	7.7%	5.6%
Level 5	31.6%	25.1%	20.5%	16.1%	15.5%	

Figure 4: Composite Score Proficiency Gain in 0.2 Increments-6-8 Grade Band

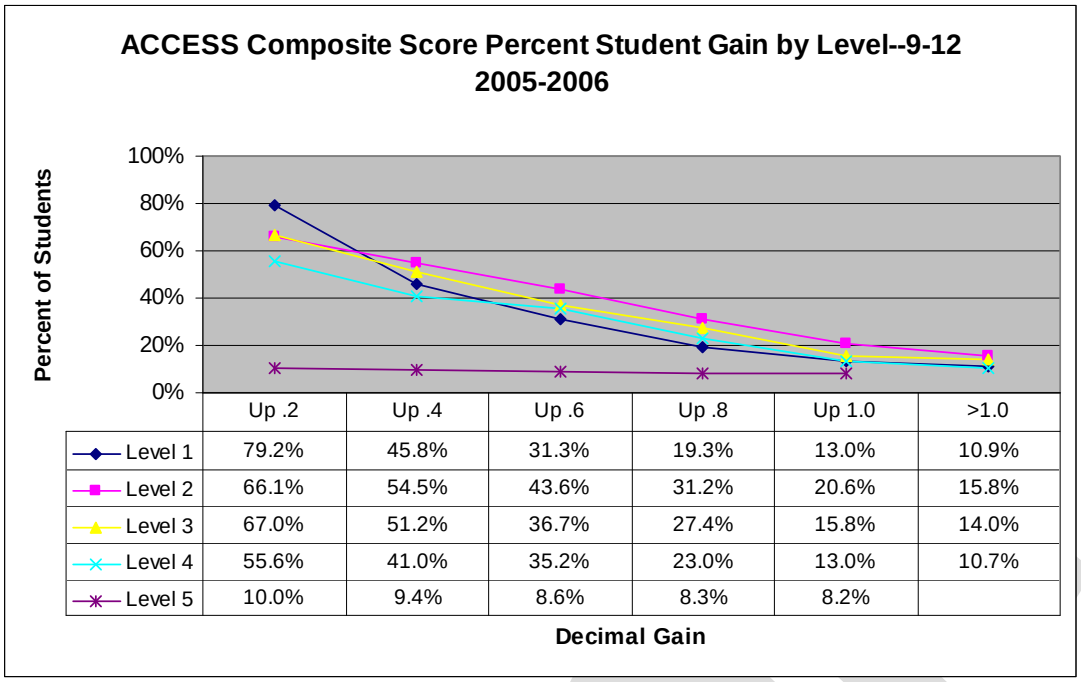


	Up .2	Up .4	Up .6	Up .8	Up 1.0	>1.0
Level 1	59.0%	40.6%	29.1%	22.2%	15.0%	11.5%
Level 2	67.8%	52.7%	39.9%	28.1%	19.9%	15.6%
Level 3	57.3%	46.0%	35.2%	24.0%	17.2%	13.3%
Level 4	48.2%	31.5%	25.1%	16.6%	10.6%	7.5%
Level 5	45.5%	37.8%	32.1%	27.4%	26.4%	

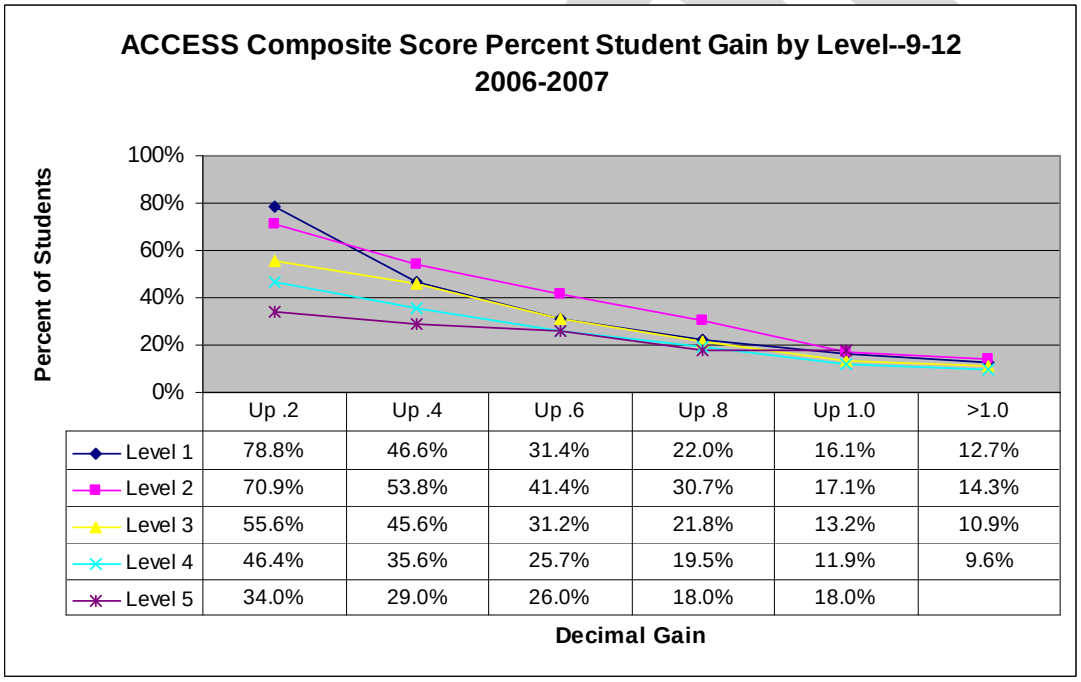


	Up .2	Up .4	Up .6	Up .8	Up 1.0	>1.0
Level 1	77.8%	59.3%	37.0%	26.9%	16.7%	13.0%
Level 2	69.2%	56.1%	39.8%	26.3%	16.6%	15.9%
Level 3	56.1%	43.8%	29.5%	20.4%	13.2%	10.7%
Level 4	42.4%	28.6%	20.7%	13.0%	8.0%	6.5%
Level 5	34.9%	25.2%	17.0%	11.5%	11.0%	

Figure 5: Composite Score Proficiency Gain in 0.2 Increments-9-12 Grade Band



	Up .2	Up .4	Up .6	Up .8	Up 1.0	>1.0
Level 1	79.2%	45.8%	31.3%	19.3%	13.0%	10.9%
Level 2	66.1%	54.5%	43.6%	31.2%	20.6%	15.8%
Level 3	67.0%	51.2%	36.7%	27.4%	15.8%	14.0%
Level 4	55.6%	41.0%	35.2%	23.0%	13.0%	10.7%
Level 5	10.0%	9.4%	8.6%	8.3%	8.2%	



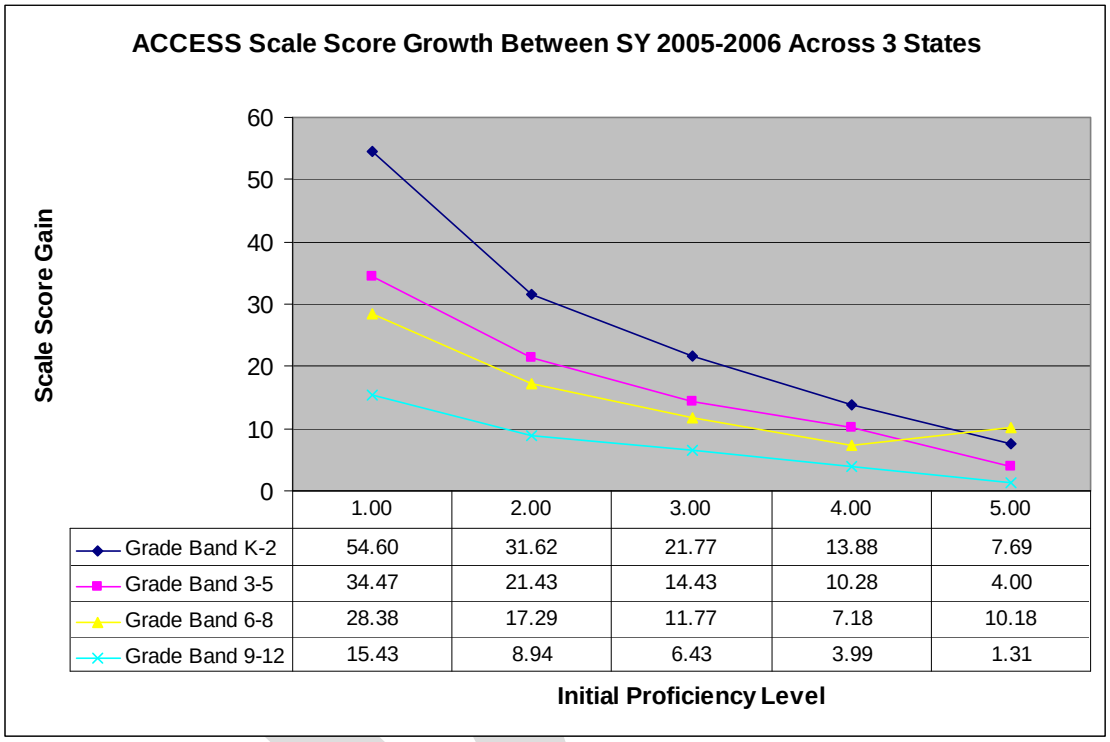
	Up .2	Up .4	Up .6	Up .8	Up 1.0	>1.0
Level 1	78.8%	46.6%	31.4%	22.0%	16.1%	12.7%
Level 2	70.9%	53.8%	41.4%	30.7%	17.1%	14.3%
Level 3	55.6%	45.6%	31.2%	21.8%	13.2%	10.9%
Level 4	46.4%	35.6%	25.7%	19.5%	11.9%	9.6%
Level 5	34.0%	29.0%	26.0%	18.0%	18.0%	

Not surprisingly, a similar trend occurs with proficiency level decimals. As initial proficiency level increases the percent of students obtaining higher scores decrease. As with the proficiency levels, findings suggest that setting different expectations based on students' proficiency level and grade band may be in order.

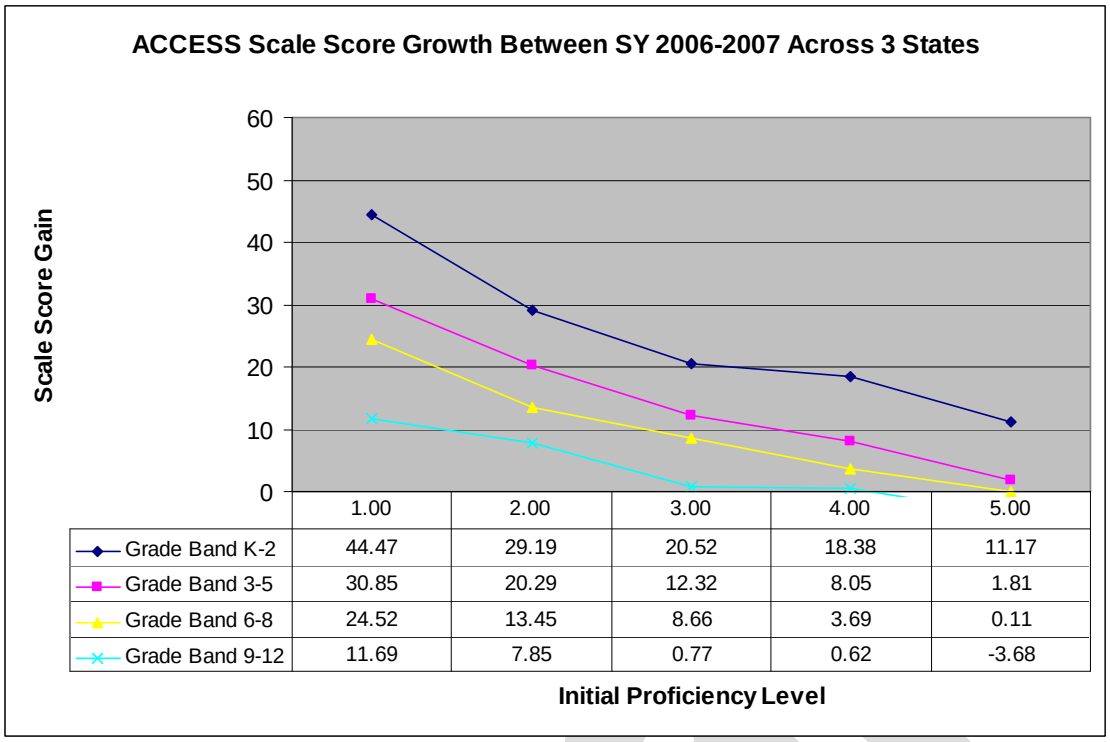
Scale Score Gain

The following graphs show mean scale score gain by grade band and starting proficiency level. Gain scores were calculated by subtracting the 2006 scale scores from the 2005 scale scores and the 2007 scores from 2006.

Figure 6: Average Scale Score Gain by Grade Band and Proficiency Level



	1.00	2.00	3.00	4.00	5.00
Grade Band K-2	54.60	31.62	21.77	13.88	7.69
Grade Band 3-5	34.47	21.43	14.43	10.28	4.00
Grade Band 6-8	28.38	17.29	11.77	7.18	10.18
Grade Band 9-12	15.43	8.94	6.43	3.99	1.31



	1.00	2.00	3.00	4.00	5.00
Grade Band K-2	44.47	29.19	20.52	18.38	11.17
Grade Band 3-5	30.85	20.29	12.32	8.05	1.81
Grade Band 6-8	24.52	13.45	8.66	3.69	0.11
Grade Band 9-12	11.69	7.85	0.77	0.62	-3.68

Notice in the 2006-2007 graph/table, higher proficiency levels in the 6-8 and 9-12 grade bands, have very small, if not negative, scale score gains. Recall that these data are within student cohorts. Students who exit ELL programs, particularly at higher levels are omitted from the 2006-2007 sample. Only students with ACCESS scores for all three years remain. It is very likely that these students have special language development issues. Thus we would expect annual scale score gain to be more reflective of the 2005-2006 graph. The trend portrayed by these data is similar to that found with proficiency levels and decimals. Students at lower grade bands and proficiency levels have greater annual scale score gains when compared to their higher grade or proficiency level peers.

Across all metrics the following patterns emerge:

- Students at lower proficiency levels grow faster
- Students in lower grades grow faster

- There is an interaction in growth between proficiency levels and grades

This finding suggests setting annual growth targets that differ based on students' initial grade and proficiency level regardless of metric used.

Determine the Starting Point and Ending Points for AMAO 1 Targets

A variety of techniques could be applied to establish starting and ending points for AMAO 1 targets. Linquanti & George used the Title I method for establishing AYP as the starting point for California's AMAO I criteria. Title I legislation directs states to establish AYP starting levels as:

> …the school at the 20th percentile in the State based on enrollment, among all schools ranked by the percentage of students at the proficient level, NCLB §1111(b)(2)(E)(ii).

Linquanti & George suggested three end points for AMAO 1: the 60th, 75th and 90th percentiles. After deliberation, the 75th percentile was chosen as California's AMAO 1 end point. We suggest setting AMAO I targets in a similar fashion. That is, the initial AMAO 1 criterion is the percent/score at which the school at the 20th percentile resides. The ending point should be discussed by expert state ELL stakeholders. These stakeholders should also have ranked school data (at relevant rankings, e.g., 60th, 75th, and 90th percentiles) to assist in setting ending points. For this report we will adopt the 20th and 75th percentiles.

States must also adopt minimum cell sizes for establishing AMAO criteria. In California's case that number was 25, i.e., at least 25 ELL students must be enrolled in a district and have two data point to be in AMAO rankings. Since different growth expectations for different proficiency levels and grade bands are being recommended, different cell size criteria are adopted for this report. If a district has 5 or more students in a cell, they are included in district rankings. Using this cell size, most districts included in rankings have 20 or more enrolled ELL students. We also adopt the most recent year of data (2006-2007) to rank districts.

Proficiency Level Starting and Ending Points

Table 4: Rankings of the Percent of Students at Varying Percentile Points by District Who Increased One or More Proficiency Levels between 2006 and 2007 School Years

Grade Band	Levels	P20	P25	P50	P75	P90
K-2	1-2	58.3	60.0	69.8	81.8	91.8
	3	40.0	41.0	52.0	70.0	75.0
	4	16.7	16.7	36.6	50.0	62.5
	5	14.1	18.2	20.0	33.3	40.0
3-5	1-2	50.0	50.0	66.7	80.0	87.5
	3	31.3	33.3	42.9	52.6	60.0
	4	12.5	14.3	20.0	33.3	45.3
	5	14.3	14.3	16.7	25.0	33.3
6-8	1-2	38.9	40.0	48.8	60.0	80.0
	3	21.4	22.7	30.0	40.0	50.0
	4	12.5	13.3	17.2	28.1	52.1
	5	14.3	15.5	18.3	22.5	25.0
9-12	1-2	22.1	28.7	42.9	58.5	77.4
	3	20.0	20.0	27.3	40.0	40.0
	4	16.7	16.7	20.0	33.3	35.3
	5	25.0	25.0	25.0	25.0	25.0

Table 4 displays the percentage of students in districts gaining one or more proficiency levels ranked by 20th, 25th, 50th, 75th, and 90th percentiles. Rankings and percentages are

arrayed by grade band and proficiency level. Note that levels 1 and 2 are collapsed; analyses suggest that this is a reasonable strategy.

Table 5: Starting and Ending Points for AMAO 1 Using the Percent of Students between Percentile Points 20 and 75 Annually Increasing One Proficiency Level or More

Grade Band	Levels	Starting Point (P20)	Smoothed Starting Point	Ending Point (P75)	Smoothed Ending Point
K-2	1-2	58.3	60	81.8	80
	3	40.0	40	70.0	70
	4	16.7	20	50.0	50
	5	14.1	15	33.3	35
3-5	1-2	50.0	50	80.0	80
	3	31.3	30	52.6	55
	4	12.5	15	33.3	35
	5	14.3	15	25.0	25
6-8	1-2	38.9	40	60.0	60
	3	21.4	20	40.0	40
	4	12.5	15	28.1	30
	5	14.3	15	22.5	25
9-12	1-2	22.1	25	58.5	60
	3	20.0	20	40.0	40
	4	16.7	15	33.3	30
	5*	25.0*	15*	25.0*	25

*Too few districts had sufficient numbers of students at this grade and level.

Table 5 outlines the AMAO 1 starting and ending points for proficiency levels. It may be desired to smooth data to communicate AMAO 1 criteria more clearly. Also, the smoothing process assures that higher grades or proficiency levels do not have more stringent criteria. For example, the district at the 75th percentile for level 4 in the 6-8 grade band is 28.1%, i.e., 28.1% of district's students in this cell gained one or more proficiency levels. The same level at the 9-12 grade band is 33.3%, which is higher. If "lower is faster and higher is slower" reflects real student growth, observed differences must be artifacts of students included in this sample. We presume this to be so and adjust

starting and ending points accordingly. Generally, smoothing is done by rounding up to the nearest 5. Certainly other strategies could be adopted.

Proficiency Level Decimal Starting and Ending Points

Table 6 presents percentile rankings for the average decimal gain by district for each grade band and proficiency level.

Table 6: Rankings of the Average Decimal Increase of Students at Varying Percentile Points by Districts between 2006 and 2007 School Years

Band	Levels	P20	P25	P50	P75	P90
K2	1-2	0.58	0.60	0.76	1.04	1.12
	3	0.21	0.32	0.44	0.62	0.87
	4	-0.02	0.03	0.30	0.44	0.72
	5	-0.62	-0.33	-0.19	0.08	0.32
3-5	1-2	0.38	0.42	0.61	0.74	0.92
	3	0.12	0.14	0.34	0.47	0.58
	4	-0.31	-0.26	-0.02	0.20	0.40
	5	-0.46	-0.44	-0.17	-0.04	0.26
6-8	1-2	0.24	0.29	0.40	0.51	0.70
	3	0.02	0.04	0.17	0.33	0.42
	4	-0.35	-0.29	-0.16	0.03	0.46
	5	-0.25	-0.24	-0.17	-0.09	0.01
9-12	1-2	-0.05	0.00	0.20	0.50	0.77
	3	-0.10	-0.07	0.14	0.31	0.60
	4	-0.43	-0.39	-0.03	0.28	0.73
	5	-0.73	-0.60	-0.30	0.10	0.40

Data arrayed in Table 6 show that rankings at lower grades and proficiency levels have higher gains, as expected. However, it should be noted that students at higher proficiency levels (especially from districts at the lowest percentiles) experience on average a decrease in proficiency level decimal units. For instance, students from districts at the 20th percentile for the 9-12 grade band, at proficiency level 5, decreased by 0.73

proficiency level decimal units between the 2006 and 2007 school years. This negative growth in decimals scores may occur for a variety of reasons:

(a) Selectivity of the sample. Students who exit from ELL programs do not take ACCESS and hence are excluded from this sample. For this reason, those students at higher proficiency levels who stay longer than expected in an ELL program may constitute a subpopulation facing particular problems (cognitive, family-related, acquisition, weak primary language literacy skills among others).

(b) Ceiling effects. Students at higher proficiency levels do not have too much room to grow as they are approaching the ACCESS test's upper-bound (proficiency level 6).

(c) Unique characteristics of WIDA proficiency decimal scores. Recall that decimals scores are not necessarily equal intervals across proficiency levels. That is, a score difference between 2.2 and 3.2 is not the same as that between 3.2 and 4.2.

(d) District-level factors. While students at higher proficiency levels from districts at the higher percentiles (for instance, percentile 90th) still make some progress, this is not the case for students from districts at lower percentiles (for instance, percentile 25th). This suggests the existence of district-level, contextual factors affecting ELL's performance.

(e) In 2006, caps (or maximum levels of English language proficiency) were placed on ACCESS tiers. A student taking a Tier A test could not exceed English language proficiency level 4 in Listening and Reading; likewise, a student taking a Tier B test, could not exceed English language proficiency level 5 in those language domains.

Taking into account these factors (especially the selectivity of our sample), we believe that for the purposes of setting starting and ending points for AMAO I targets smoothing is necessary. That is to say, we do not expect students to decrease in their language learning progress. Rather, we should expect them to make some gain even if modest, and this should also be the expectation at the district level. Table 7 shows the results of smoothing using proficiency decimal scores, at the 20th and 75th percentile, for each grade level cluster.

Table 7: Starting and Ending Percentile Points for AMAO 1 Using Proficiency Decimal Scores

Grade Band	Levels	Starting Point (P20)	Smoothed Starting Point	Ending Point (P75)	Smoothed Ending Point
K-2	1-2	0.58	0.60	1.04	1.00
	3	0.21	0.30	0.62	0.80
	4	-0.02	0.20	0.44	0.60
	5	-0.62	0.10	0.08	0.40
3-5	1-2	0.38	0.40	0.74	0.80
	3	0.12	0.20	0.47	0.60
	4	-0.31	0.10	0.20	0.40
	5	-0.46	0.10	-0.04	0.20
6-8	1-2	0.24	0.30	0.51	0.60
	3	0.02	0.20	0.33	0.40
	4	-0.35	0.10	0.03	0.30
	5	-0.25	0.10	-0.09	0.20
9-12	1-2	-0.05	0.20	0.50	0.50
	3	-0.10	0.10	0.31	0.40
	4	-0.43	0.10	0.28	0.30
	5	-0.73	0.10	0.10	0.20

Scale Score Starting and Ending Points

Table 8 shows district rankings of the average scale score increase between the 2006 and 2007 school years for the three state samples.

Table 8: Rankings of the Average Scale Score Increase of Students at Varying Percentile Points by Districts between 2006 and 2007 School Years

Band	Levels	P20	P25	P50	P75	P90
K2	1-2	28	28	33	42	43
	3	18	18	21	25	32
	4	12	13	18	21	27
	5	0	6	11	20	25
3-5	1-2	21	24	26	33	45
	3	12	13	17	20	23
	4	3	5	9	14	19
	5	1	2	10	12	15
6-8	1-2	18	18	21	25	33
	3	8	9	11	14	17
	4	0	1	4	8	16
	5	3	3	4	6	9
9-12	1-2	1	4	11	20	30
	3	1	2	8	11	19
	4	-5	-3	3	9	18
	5	-9	-7	-2	8	14

At higher grade levels, findings similar to decimal gains are observed, i.e., higher grades and proficiency levels have low or negative gains. Potential reasons for this low gain have been stated earlier. Again, smoothing is necessary to provide meaningful AMAO 1 expectations with scale scores. Table 9 displays smoothed scale scores for starting and ending points.

Table 9: Starting and Ending Points for AMAO 1 Using Scale Scores

Grade Band	Levels	Starting Point (P20)	Smoothed Starting Point	Ending Point (P75)	Smoothed Ending Point
K-2	1-2	28	28	42	42
	3	18	18	25	25
	4	12	12	21	21
	5	0	2	20	20
3-5	1-2	21	21	33	33
	3	12	12	20	20
	4	3	3	14	14
	5	1	2	12	12
6-8	1-2	18	18	25	25
	3	8	8	14	14
	4	0	3	8	8

	5	3	2	6	6
9-12	1-2	1	2	20	20
	3	1	2	11	11
	4	-5	2	9	9
	5	28	2	8	8

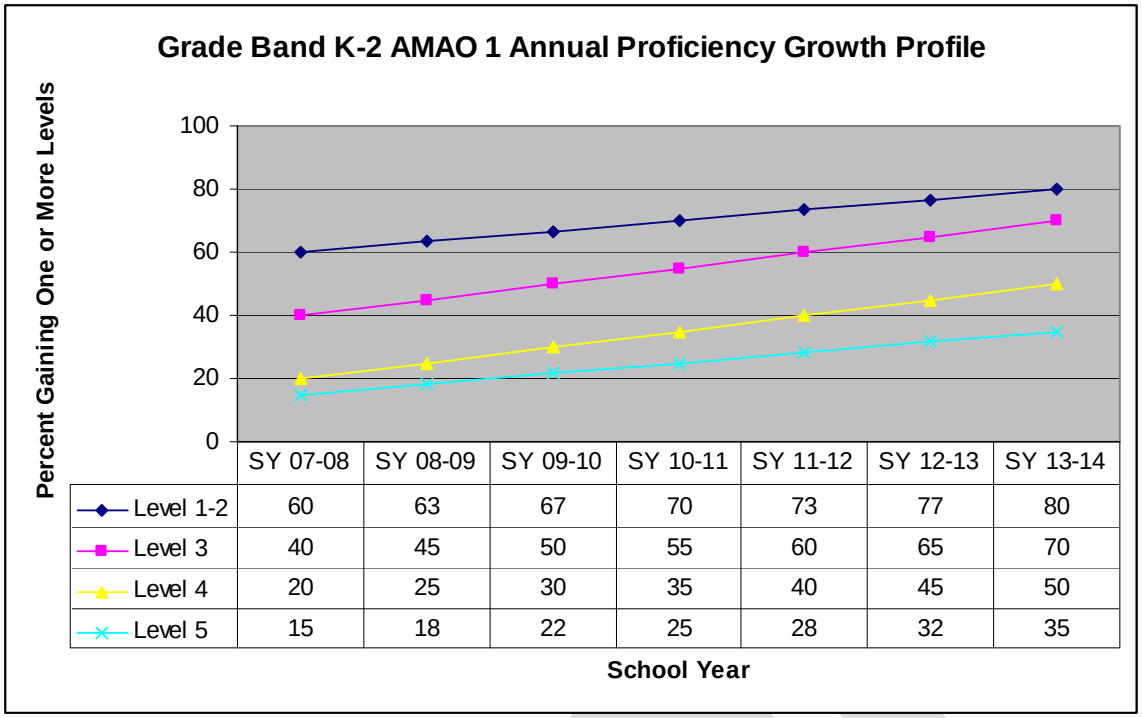
The smoothing process for scale scores focuses primarily on scores that are 1 or less. More smoothing may be desired, e.g., rounding down or up to the nearest 5.

Determine Annual Rate of Growth from 2006 – 2014

Once starting and ending points are established annual increases in district growth can be projected. The following series of figures project annual growth rates from 2006 to 2014 for proficiency levels, proficiency level decimals, and scale scores based on starting and ending points described earlier. The dates selected here coincide with NCLB deadlines for AYP. States may choose to have other projected timeframes.

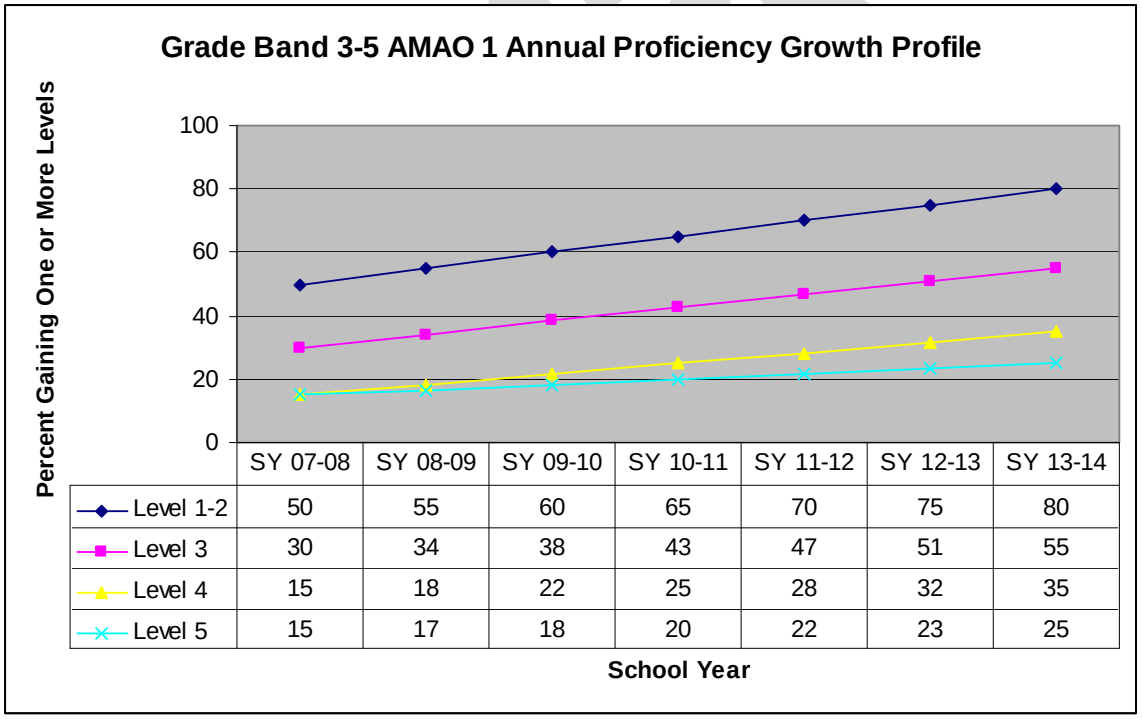
Proficiency Level AMAO 1 Profiles

Figure 7: K-2 AMAO 1 Profile for Proficiency Scores



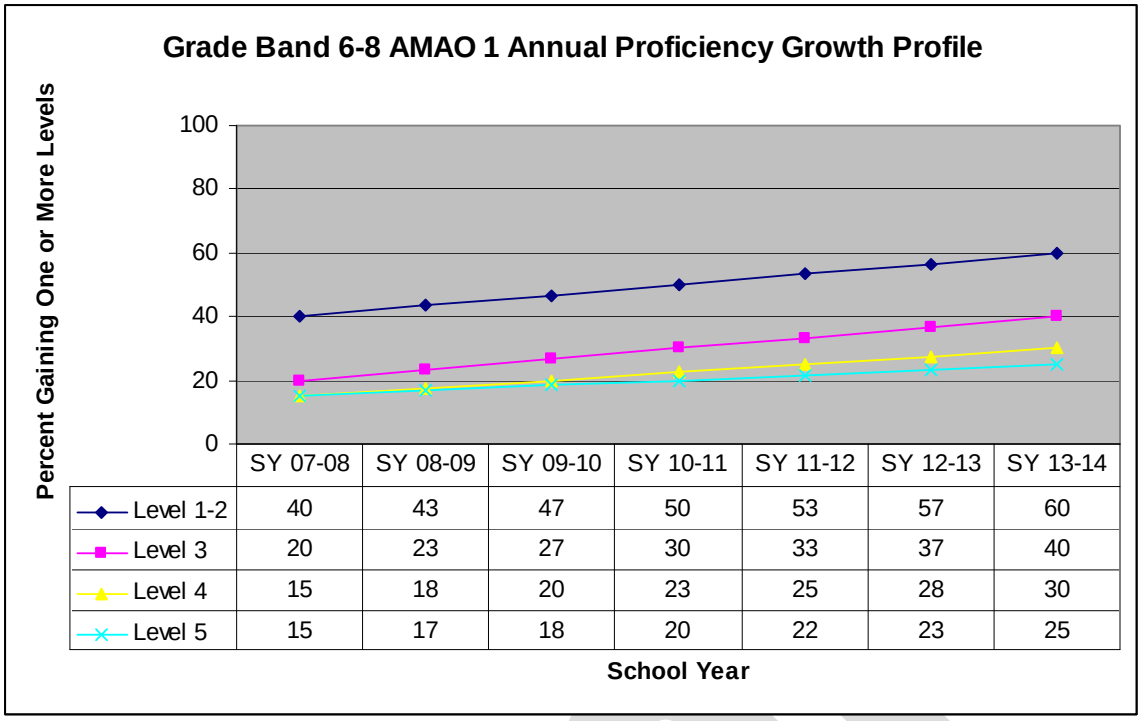
	SY 07-08	SY 08-09	SY 09-10	SY 10-11	SY 11-12	SY 12-13	SY 13-14
Level 1-2	60	63	67	70	73	77	80
Level 3	40	45	50	55	60	65	70
Level 4	20	25	30	35	40	45	50
Level 5	15	18	22	25	28	32	35

Figure 8: 3-5 AMAO 1 Profile for Proficiency Scores



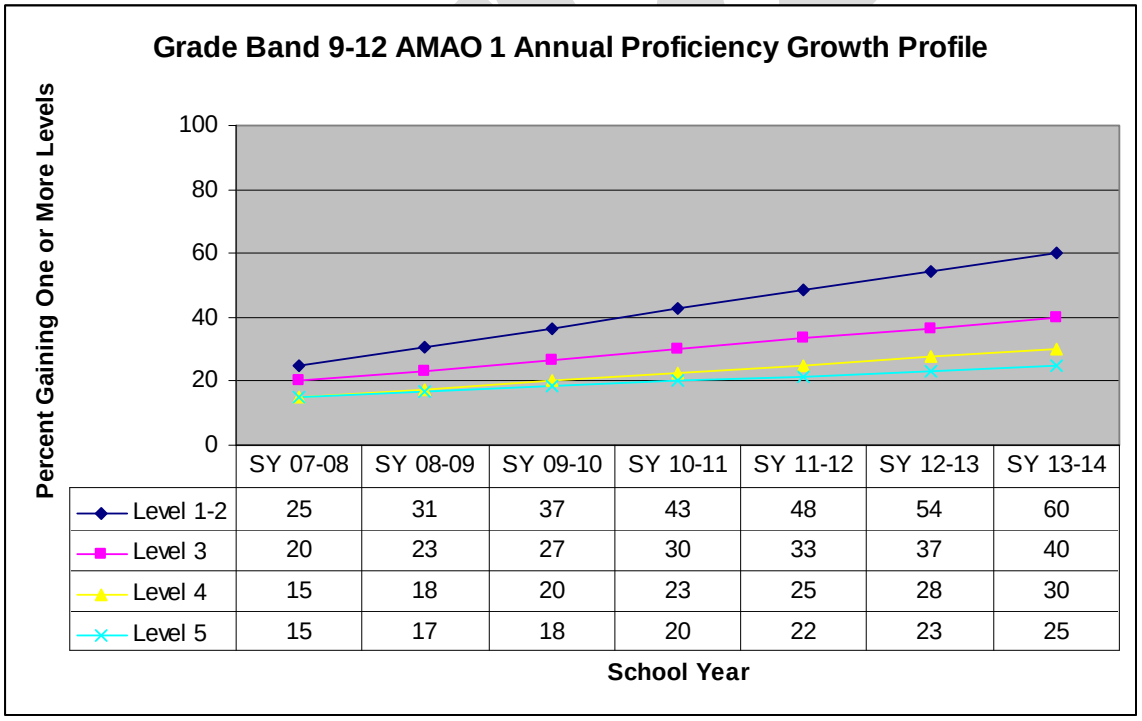
	SY 07-08	SY 08-09	SY 09-10	SY 10-11	SY 11-12	SY 12-13	SY 13-14
Level 1-2	50	55	60	65	70	75	80
Level 3	30	34	38	43	47	51	55
Level 4	15	18	22	25	28	32	35
Level 5	15	17	18	20	22	23	25

Figure 9: 6-8 AMAO 1 Profile for Proficiency Scores



	SY 07-08	SY 08-09	SY 09-10	SY 10-11	SY 11-12	SY 12-13	SY 13-14
Level 1-2	40	43	47	50	53	57	60
Level 3	20	23	27	30	33	37	40
Level 4	15	18	20	23	25	28	30
Level 5	15	17	18	20	22	23	25

Figure 10: 9-12 AMAO 1 Profile for Proficiency Scores



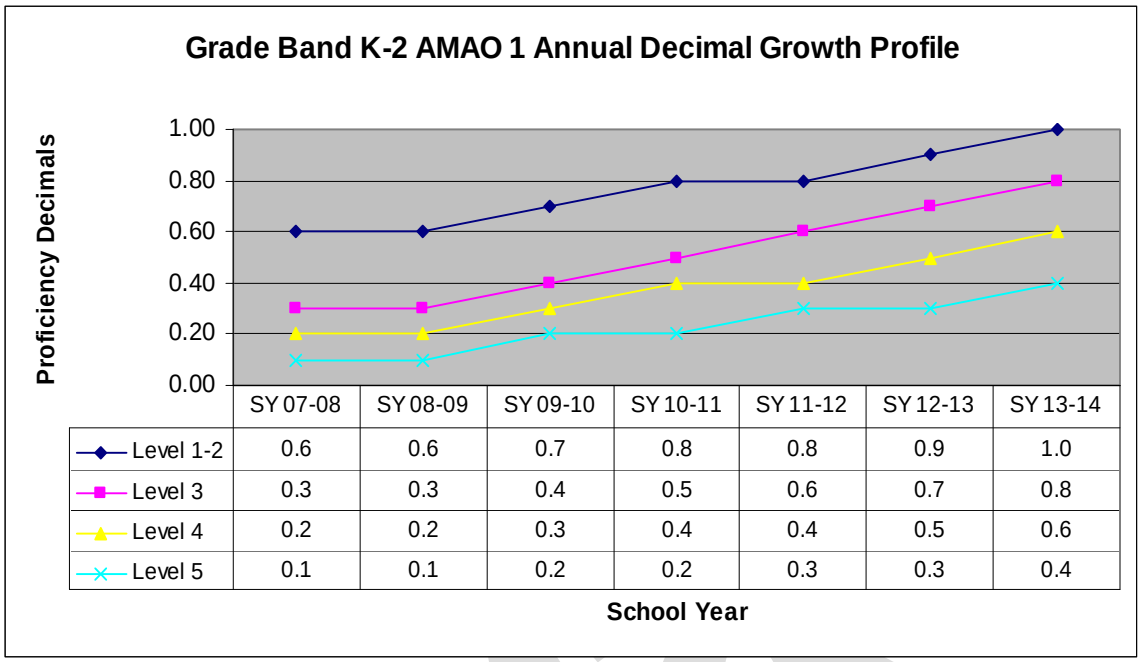
	SY 07-08	SY 08-09	SY 09-10	SY 10-11	SY 11-12	SY 12-13	SY 13-14
Level 1-2	25	31	37	43	48	54	60
Level 3	20	23	27	30	33	37	40
Level 4	15	18	20	23	25	28	30
Level 5	15	17	18	20	22	23	25

Values in cells in the figures above are the percent of students in districts who should gain one or more proficiency levels per year. AMAO 1 growth profiles in the above

figures project linear annual increases. States may opt for slower initial AMAO expectations and then ramp up growth profiles as time progresses. The proficiency level decimal AMAO 1 growth profiles adopt this strategy.

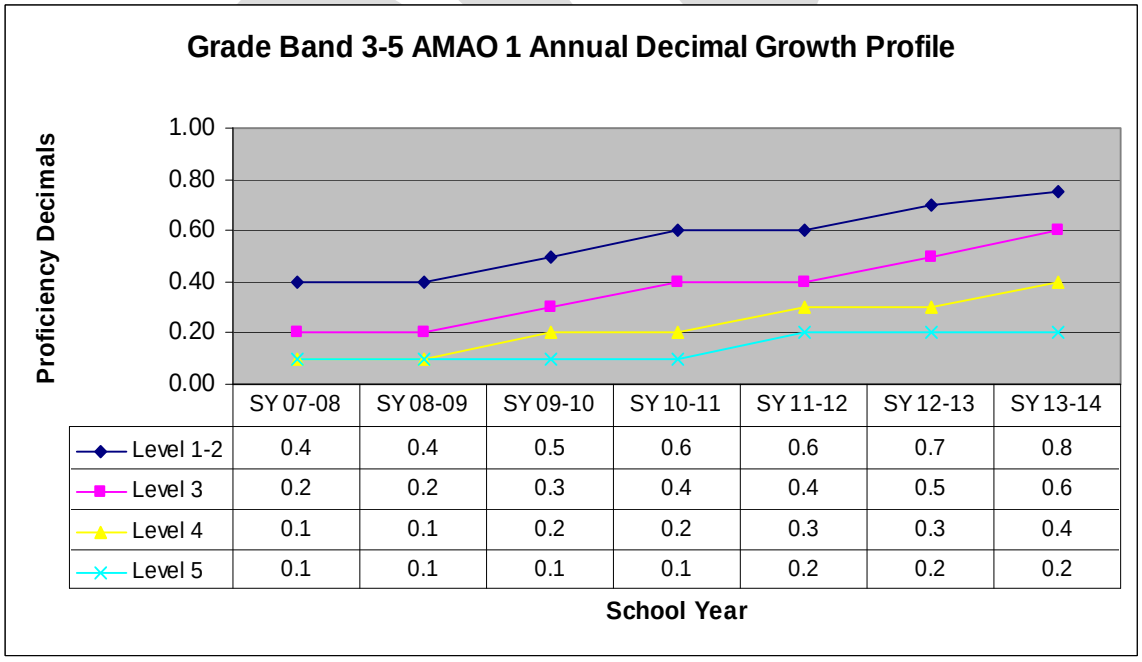
Proficiency Level Decimal AMAO 1 Profiles

Figure 11: K-2 AMAO 1 Profile for Proficiency Level Decimal Scores



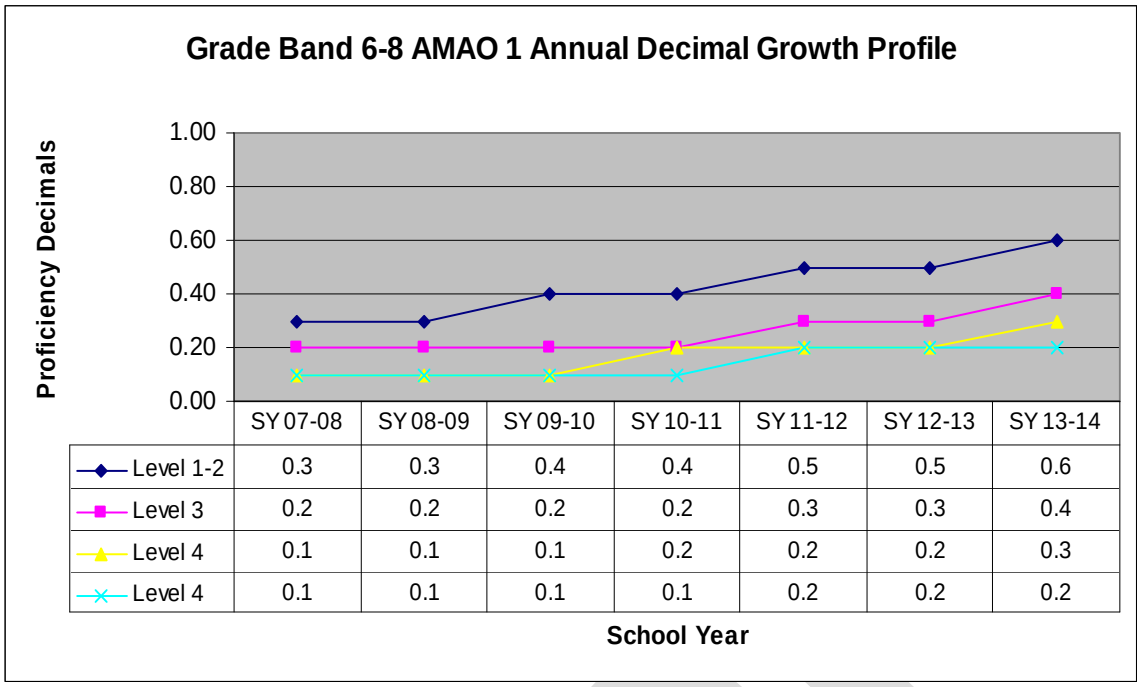
	SY 07-08	SY 08-09	SY 09-10	SY 10-11	SY 11-12	SY 12-13	SY 13-14
Level 1-2	0.6	0.6	0.7	0.8	0.8	0.9	1.0
Level 3	0.3	0.3	0.4	0.5	0.6	0.7	0.8
Level 4	0.2	0.2	0.3	0.4	0.4	0.5	0.6
Level 5	0.1	0.1	0.2	0.2	0.3	0.3	0.4

Figure 12: 3-5 AMAO 1 Profile for Proficiency Level Decimal Scores



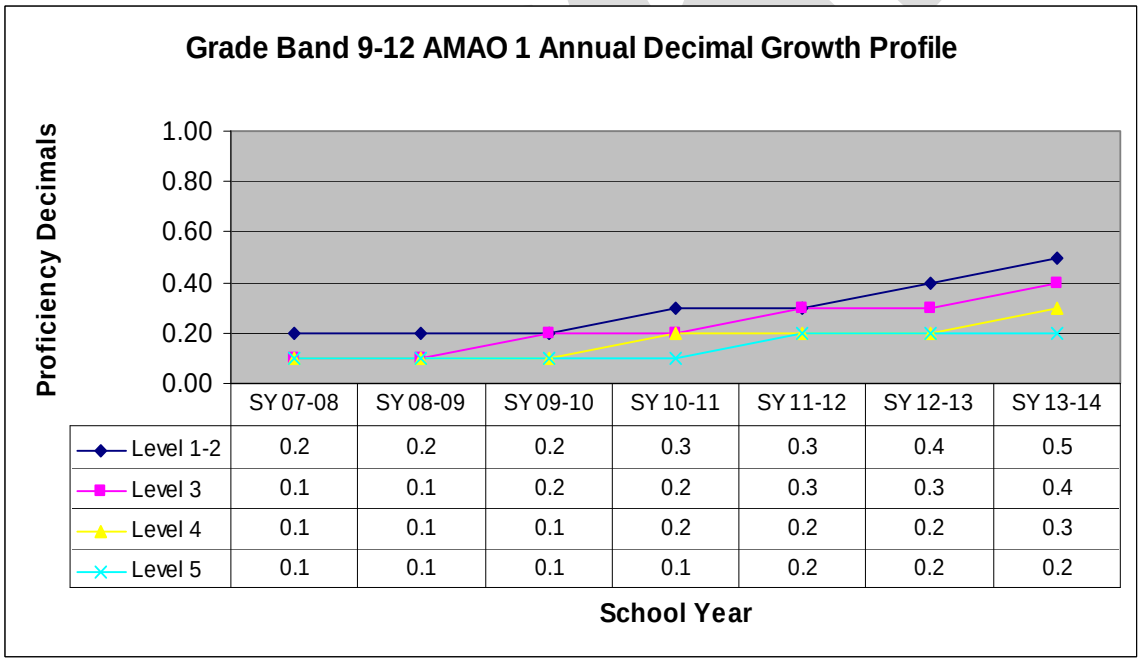
	SY 07-08	SY 08-09	SY 09-10	SY 10-11	SY 11-12	SY 12-13	SY 13-14
Level 1-2	0.4	0.4	0.5	0.6	0.6	0.7	0.8
Level 3	0.2	0.2	0.3	0.4	0.4	0.5	0.6
Level 4	0.1	0.1	0.2	0.2	0.3	0.3	0.4
Level 5	0.1	0.1	0.1	0.1	0.2	0.2	0.2

Figure 13: 6-8 AMAO 1 Profile for Proficiency Level Decimal Scores



	SY 07-08	SY 08-09	SY 09-10	SY 10-11	SY 11-12	SY 12-13	SY 13-14
Level 1-2	0.3	0.3	0.4	0.4	0.5	0.5	0.6
Level 3	0.2	0.2	0.2	0.2	0.3	0.3	0.4
Level 4	0.1	0.1	0.1	0.2	0.2	0.2	0.3
Level 4	0.1	0.1	0.1	0.1	0.2	0.2	0.2

Figure 14: 9-12 AMAO 1 Profile for Proficiency Level Decimal Scores

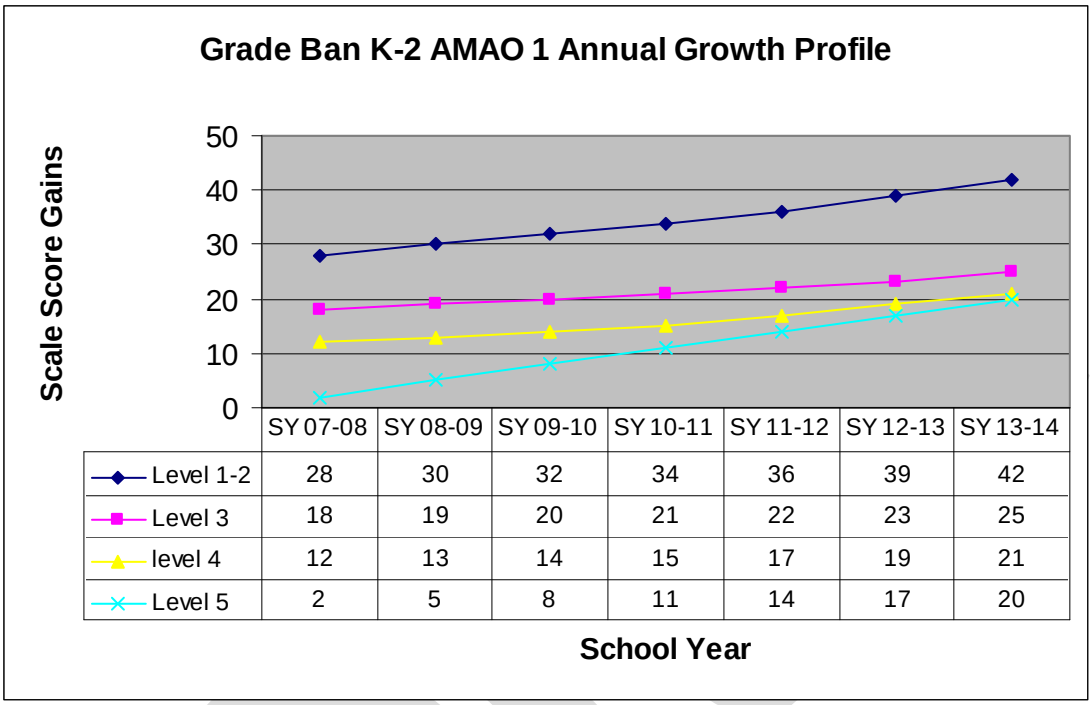


	SY 07-08	SY 08-09	SY 09-10	SY 10-11	SY 11-12	SY 12-13	SY 13-14
Level 1-2	0.2	0.2	0.2	0.3	0.3	0.4	0.5
Level 3	0.1	0.1	0.2	0.2	0.3	0.3	0.4
Level 4	0.1	0.1	0.1	0.2	0.2	0.2	0.3
Level 5	0.1	0.1	0.1	0.1	0.2	0.2	0.2

Scale Score AMAO 1 Profiles

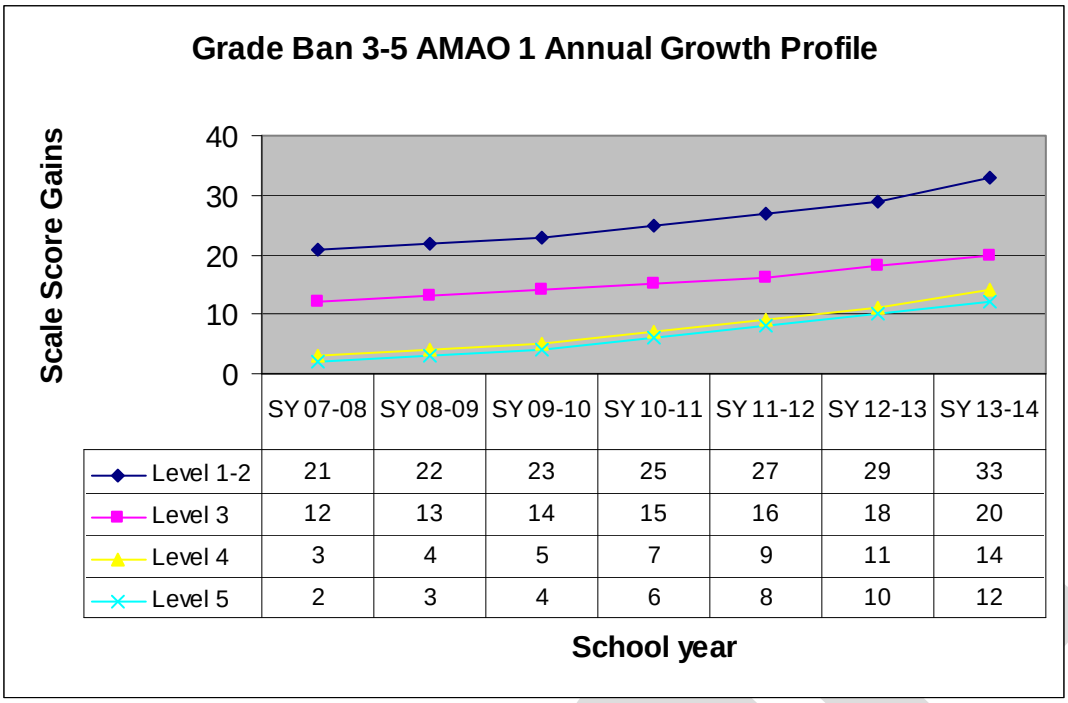
Figure 15 through Figure 18 show AMAO scale score growth profiles.

Figure 15: K-2 AMAO 1 Profile for Scale Scores



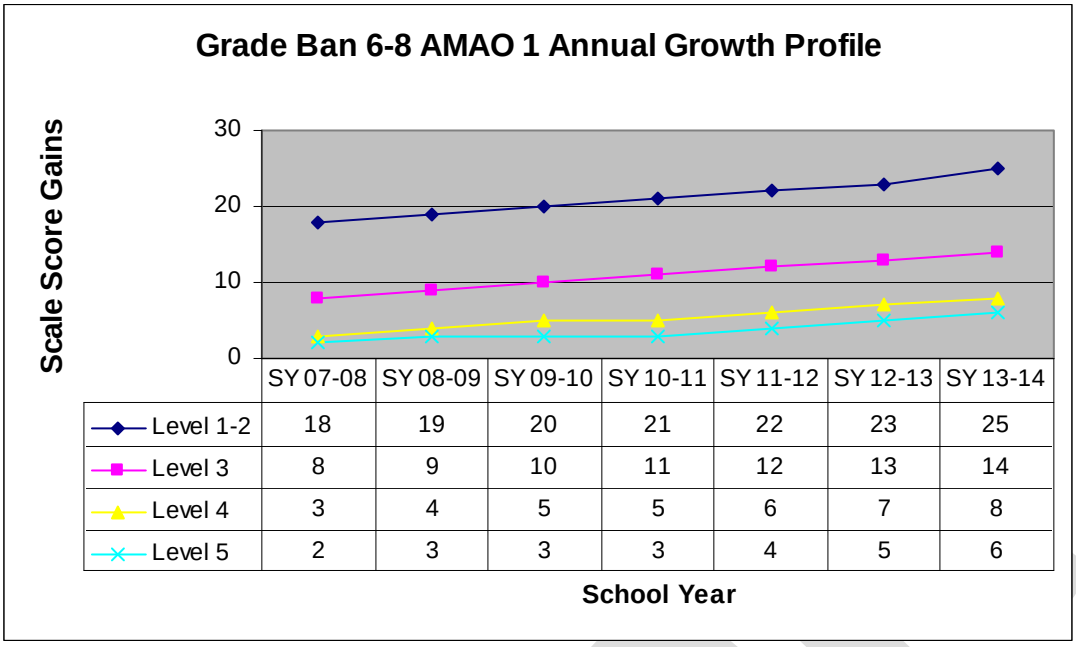
	SY 07-08	SY 08-09	SY 09-10	SY 10-11	SY 11-12	SY 12-13	SY 13-14
Level 1-2	28	30	32	34	36	39	42
Level 3	18	19	20	21	22	23	25
level 4	12	13	14	15	17	19	21
Level 5	2	5	8	11	14	17	20

Figure 16: 3-5 AMAO 1 Profile for Scale Scores



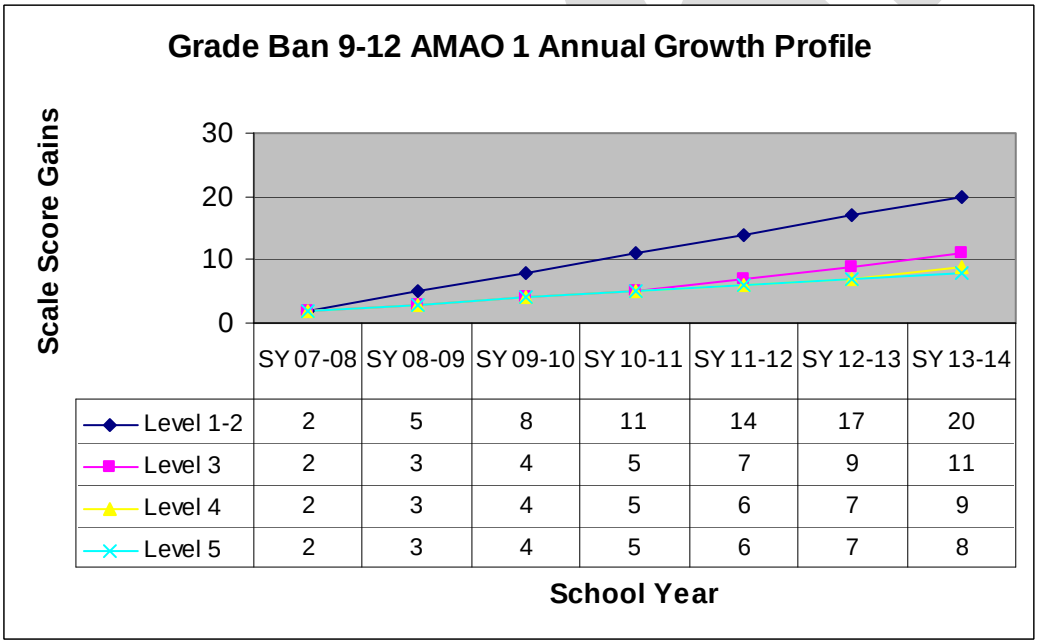
	SY 07-08	SY 08-09	SY 09-10	SY 10-11	SY 11-12	SY 12-13	SY 13-14
Level 1-2	21	22	23	25	27	29	33
Level 3	12	13	14	15	16	18	20
Level 4	3	4	5	7	9	11	14
Level 5	2	3	4	6	8	10	12

Figure 17: 6-8 AMAO 1 Profile for Scale Scores



	SY 07-08	SY 08-09	SY 09-10	SY 10-11	SY 11-12	SY 12-13	SY 13-14
Level 1-2	18	19	20	21	22	23	25
Level 3	8	9	10	11	12	13	14
Level 4	3	4	5	5	6	7	8
Level 5	2	3	3	3	4	5	6

Figure 18: 9-12 AMAO 1 Profile for Scale Scores



	SY 07-08	SY 08-09	SY 09-10	SY 10-11	SY 11-12	SY 12-13	SY 13-14
Level 1-2	2	5	8	11	14	17	20
Level 3	2	3	4	5	7	9	11
Level 4	2	3	4	5	6	7	9
Level 5	2	3	4	5	6	7	8

There are two final points to consider when setting AMAO 1 criteria. First, students' observed scores and districts' observed rankings strongly point to setting differing growth expectations at different grade bands and proficiency levels. In the tables and figures in

this section, proficiency levels and grade bands were broken into four categories each. This was done primarily for illustrative purposes. Alternative breakouts are certainly plausible, for example K-5 and 6-12 for grade bands and 0-2.5, 2.6-3.5 and ≥3.6 for proficiency levels. Examining available state data prior to adopting grade band and/or proficiency level distinctions is strongly suggested. Second, the above AMAO starting and ending points were set using district rankings. Some states may have relatively few districts, and percentile rankings may not be meaningful. In this case, examining student rankings and distributions may inform starting and ending points. But this strategy is problematic, since the variance in growth of all students at the state level is typically far greater than the aggregate variance in district growth. Setting growth expectations based students will tend to be higher than what districts are capable of. In the case of states with small numbers of districts, examining both state and district variance in growth may be the best strategy for setting starting and ending points for AMAO 1.

Development of Annual Measurable Achievement Objective 2: English Language Proficiency

Linquinti and George (in press) suggest five decisions that need to be made to establish AMAO 2 targets. They are:

1. Define the English proficient level
2. Determine the cohort of ELLs for analysis
3. Set the starting point for AMAO 2 targets
4. Set the ending point for AMAO 2 targets
5. Determine the rate of annual growth

As can be seen, AMAO 2 decision points are very similar to those of AMAO 1. The critical differences lie in the first two points, and that is where we turn to next.

Define English Language Proficiency

What is English language proficiency in the K-12 school context? This question has been a hotly debated issue in the fields of bilingual education, teaching English to speakers of other languages, and applied linguistics for years (Cummins 1983, Collier 1995, Hakuta, Butler, and Witt, 2002). Some definitions focus on students' linguistic capabilities to manage academic English contexts (e.g., Collier 1995). Others take a more pragmatic view and focus on how students perform or are predicted to perform on standardized state assessments (e.g., Hakuta, et al, 2002). Federal Law provides guidance on how students are defined as limited English proficient. That is, a limited English proficient (LEP) student is an enrolled, school aged child:

- whose native language is not English, and
- whose difficulties in English deny them the ability to perform proficient on the state's achievement test, and
- whose difficulties in English deny them the ability to successfully participate in class where only English is spoken, and
- whose English language ability deny them the opportunity to participate fully in society.

(NCLB §9101(25))

Federal guidance and research identify three characteristics of LEP students: their native language is not English, their academic performance is affected by their proficiency in

English, and their social flexibility is limited by their language proficiency. The corollary to defining an LEP student would be the definition of an English proficient student, i.e., has the ability to perform proficiently on state tests, successfully participates in classes where only English is spoken, and fully participates in English lingua franca social environments.

Based on NCLB, states have created English language proficiency standards and with these standards defined what "English Proficient" means. Most, if not all, states empanelled state (sometimes national) stakeholder experts to develop standards and English language proficiency expectations. We suggest following a similar strategy with some additions to setting AMAO 2 expectations. In our view, it is productive to establish English proficiency (as it relates to AMAOs) similarly to how state English language proficiency standards are established. (For more information on this topic see Cizek (2001) and Hambleton & Pitoniak (2006).) We suggest the following activities be considered when setting AMAO 2 English proficient expectations:

- Empanel relevant stakeholder experts to define what AMAO 2 English language proficiency means in relation to WIDA English language proficiency standards,
- Correlate the state's reading and mathematics test data to ACCESS scores and determine which performance in ACCESS is predictive of success on the reading and math assessments,
- Identify ELL students who successfully participate in class without substantial English language support and identify which ACCESS proficiency level(s) that represents.

A careful reading of the Linquanti & George report reveals that AMAO 2 expectations were set in a similar fashion. To the first activity, a possible list of relevant stakeholders might be parents, teachers, administrators, and interested parties who are familiar and have worked with ELL students as well as with the state's English language proficiency standards and expectations. Panelists from stakeholder groups could be provided information from the second two activities to deliberate upon to set AMAO 2 levels. As with standards setting activities, this activity could be conducted with several rounds until agreed upon levels are established.

Regarding the second activity, some sort of correlation or regression analysis should be considered when examining the relationship between a state's content assessments and ACCESS. For example, biserial correlations between the state's proficiency levels in reading and mathematics and ACCESS scales scores at each grade band could be conducted. One might also consider regressing ACCESS proficiency scores to the state's reading and mathematics scores. In both cases, we would be looking for the point at which ACCESS scores begin to NOT be predictive or begin to have lower correlations. This may seem counter intuitive, but consider what English proficient conceptually context means in the K-12, "the point at which a student's English language proficiency no longer interferes with their ability to fully participate in English-only curriculum and in English-only academic contexts." *If this conceptualization is the basis for defining K-12 second language English proficiency, then it follows that the point at which language limitations no longer predict academic performance would be the place to set English proficient*. Said differently, language is no longer the limiting factor in determining a

students' performance on the state's achievement tests. Federal law indicates that LEP students are defined as those who are limited in their "ability to perform proficient on the state's achievement test(s)", not necessarily those who are proficient on the state's achievement tests. ***One could argue that the point at which the distribution of ELL students on the state's achievement tests closely matches that of first language English speakers represents English proficient***. The challenge of defining proficiency thus is not to set that point too low or too high.

For the final activity, a focus group, survey, observational protocol and/or artifact analysis could be used to identify the level at which ELL students meaningfully participate in English only classrooms. Data from these analyses would be aggregated and summarized. Students' ACCESS data could be matched with aggregated data and used to set AMAO 2 levels. Information from all three activities should be used in concert when defining English proficiency as it relates to AMAO 2 expectations. Use of all available information and data sources will provide meaningful, attainable and challenging expectations.

Determine the Cohort of ELLs for Analysis

Linquanti & George (in press) write,

> NCLB Title III requires that AMAOs be developed in a manner that reflects the amount of time an individual child has been enrolled in a language instruction educational program. This AMAO there entails a cohort analysis. One key issue is *which ELL students can reasonably be expected to reach English language proficiency at a given point in time*…. (p.7)

The question behind cohort identification is, "How long should it take for a student to become English proficient?" There are two potential methods to address this question: estimate the time a student at a particular grade-band and proficiency level would take to become English proficient or identify students who would be included in a cohort from which English proficient percentages would be calculated.

Approach 1

If you determine cohorts based on time, you must first statistically model the time required to be English proficient based on a students' starting grade-band and proficiency level. AMAO 1 analyses of ACCESS data across three states revealed that growth in language proficiency over time was not linear but curvilinear, such that younger students and lower proficiency levels grew faster than older students and higher proficiency levels, but the rate of growth declined as students moved up proficiency levels. Thus non-linear models would need to be fitted.

Next, you would create a matrix of expected times based on a students' grade-band and proficiency levels. Once done, you would need to track students over time keeping in mind where they started and where they currently are relative to their grade and proficiency level. It would then be necessary to determine the percent of students who would be "on-track" in districts. This process is appealing in that it is consistent with current understanding of child second language acquisition and in concert with observations. However, several challenges arise when applying this method. First, non-linear models would need to be fitted to meaningfully predict student growth. This requires sophisticated statistical analyses that state universities have the capacity to

perform. However, it is unclear if state departments of education, especially Title III directors and their staff, have the capacity to apply these models, not to mention the background to describe them to districts, schools and teachers. Another challenge is the amount of record keeping needed to manage such a system. A state longitudinal student identification system would be a must for this method. It would be necessary to track what grades and proficiency levels students began their English language instructional program, where they are currently at in their programs, and when they exited. Certainly most districts have this capacity, but this would be a requirement at the state level. This approach may prove challenging for many states.

Approach 2

An alternative would be to set AMAO 2 cohorts based on students at different proficiency levels. This is the approach Linquanti & George take and the approach we suggest. To apply this method, you first need to assign cohorts of students included in the proficiency calculation. Linquanti & George write that the California state board decided on the following criteria (Note California's proficiency levels are Beginning, Early Intermediate, Intermediate, Early Advanced, and Advanced.):

- ELLs who were at the Intermediate level in the prior year
- ELLs who were at Early Advanced and Advanced but not English proficient in the prior year
- ELLs who were at Beginning or Early Intermediate in the prior year and who first enrolled in US schools four or more years ago

- ELLs who were at Beginning or Early Intermediate in the prior year and in US schools *less than* four years who reach the English proficient level in the current year. (p.8)

Two important questions were answered in forming these cohort criteria. First, at what level should students be included into the English proficient formula for districts? In California's case, that level was below Intermediate. Second, how long should students in the lowest levels be excluded from the English proficient formula? California decided 4 years was the appropriate time frame. California added another criterion; students who were in the lowest levels (normally excluded) but reached proficiency should also be included. In determining cohort assignment using ACCESS data and WIDA proficiency levels, we suggest using the same procedure.

In examining student proficiency level distributions in the three state sample and running a variety of cohort combinations, a WIDA proficiency level of 2.5 seemed to provide similar percentages as that identified in Linquanti and George's paper. Further, over a three year period in the three state sample, less than 3% of student who started with a 2.5 or less in 2005 received a level 5.0 or higher in 2007. It is very unlikely that students would be able to meet the English proficient time frame in 3 years. In this case, as with Linquanti and George, a 4-year time frame seems reasonable. Given this, we would have the following cohort groups:

- ELLs who were at a WIDA proficiency level of 2.5 or higher in the previous year,
- ELLs who were at a WIDA proficiency level of less than 2.5 and who first enrolled in US schools four or more years ago,

- ELLs who were at a WIDA proficiency level of less than 2.5 in the prior year and in US schools *less than* four years who reach the English proficient level in the current year.

As stated earlier, each state must address the questions of when should students start being counted and how long should lower level students be excluded from the English proficient calculation. The cohort criteria above should be a good place to begin discussing theses issues.

Determine the Starting Point

To establish a starting point, Linquanti & George applied California's AMAO 2 cohort criteria to districts and rank ordered districts by the percent of students attaining English proficiency for those having 25 or more ELL students. The starting point was determined to be the percent English proficient of district at the 20th percentile. For illustrative purposes, we will adopt a WIDA proficiency level of 5.0 as the English proficient point. Further, we will perform the same rank ordering as Linquanti & George, but we will set the minimum number of students to 20 within a district instead of 25. For this analysis, we used the 2005-2006 data to rank districts. When all districts in the three state sample were rank ordered, the district at the 20th percentile corresponded to 18 percent proficient. This means that 18% of all students in the AMAO 2 cohort were expected to reach proficiency by the 2005-2006 school year.

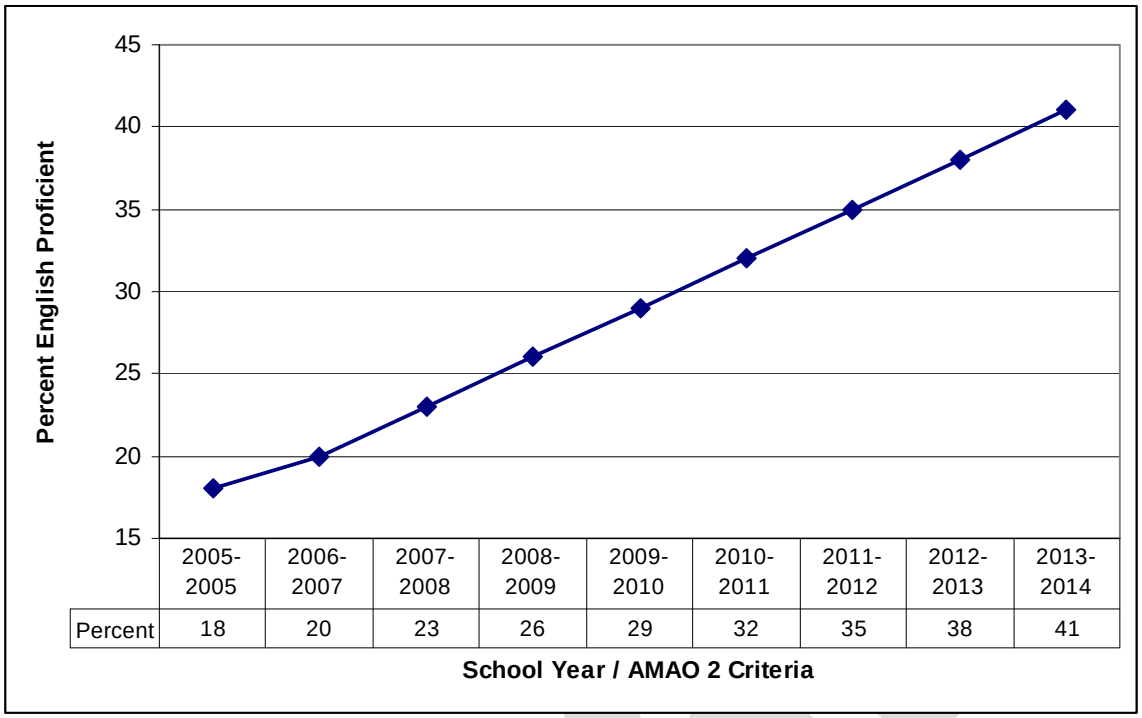
Determine the End Point

Federal law requires schools to annually increase the numbers of students attaining English proficiency. Fortunately, Title III does no require the end point of English proficiency to be 100%, which would not be possible. Where should the end point be? This is certainly something that should be discussed. Three alternatives were provided to California policymakers: the 60th, 75th and 90th percentiles. California chose the 75th percentile, which corresponded to the 46% English proficient for AMAO 2. Using the WIDA three state dataset, the 75th percentile would correspond to 41% English proficient.

Determine the Rate of Annual Growth from 2005 to 2014

With starting and ending points obtained, an annual growth rate can be calculated. Assuming a WIDA level of 5.0 as English proficient, which would be a starting point of 18% and an ending point of 41%. The figure below plots a 9-year timeline of annual growth to 2014. Again this figure is for illustrative purposes, each state may chose to have a different objectives and timelines.

Figure 19: Example of AMAO 2 Targets



With 2005-06 as a starting point, districts would be required to have 23% or more of their students in cohort groups reach English proficiency for the 2007-2008 school year.

One concern with the AMAO 2 targets presented above might the difference in the percent of English proficient students at particular grades. If younger students grow faster, districts that service predominantly older (e.g., middle school or high school) students may be disadvantaged by having one criterion across all grade bands. In many WIDA member states, there are "unified" or high school districts that serve only high school students. There may be a need to create grade-band based AMAO 2 targets. For example, one might create K-5 and 6-12 AMAO 2 starting and ending points as well as annual growth expectations. This was done with the current three state dataset and little difference was found between the K-5 or 6-12 grade band groups. Nonetheless, we would encourage states to examine this possibility.

Summary

The primary purpose of this paper is to provide guidance to WIDA states in setting Annual Measurable Achievement Objectives (AMAOs) in English as required under No Child Left Behind legislation. We also hope that this paper promotes further dialogue within WIDA states and beyond about the complexities involved with setting challenging yet reasonable growth targets for this diverse group of students we refer to as English language learners.

AMAO 1-Findings Recommendations

- There are three possible metrics available for setting AMAO 1 criteria with the ACCESS Assessment: proficiency levels, proficiency level decimals, and scale scores.
- For all metrics, there was different progress in students' language development (based on ACCESS scores) at different grades and proficiency levels, such that student at lower grades/proficiency levels progressed faster than students at higher grades/proficiency levels. This suggested the need to set different growth expectations based on initial grade and proficiency level.

The tables below summarize key AMAO 1 findings based on the three-state WIDA dataset.

Table 10: AMAO 1 Starting and Ending Points based on Number of Students Gaining One or More Proficiency Levels Per Year

Grade Band	Percent Gain by Initial Proficiency Level				
	Criteria	1-2	3	4	5
K-2	Start	60%	40%	20%	15%
	End	80%	70%	50%	35%
3-5	Start	50%	30%	15%	15%
	End	80%	55%	35%	25%
6-8	Start	40%	20%	15%	15%
	End	60%	40%	30%	25%
9-12	Start	25%	20%	15%	15%
	End	60%	40%	30%	25%

Table 11: AMAO 1 Starting and Ending Points based on Annual Proficiency Level Decimal Score Increase

Grade Band	Percent Gain by Initial Proficiency Level				
	Criteria	1-2	3	4	5
K-2	Start	0.60	0.30	0.20	0.10
	End	1.00	0.80	0.60	0.40
3-5	Start	0.40	0.20	0.10	0.10
	End	0.80	0.60	0.40	0.20
6-8	Start	0.30	0.20	0.10	0.10
	End	0.60	0.40	0.30	0.20
9-12	Start	0.20	0.10	0.10	0.10
	End	0.50	0.40	0.30	0.20

Table 12: AMAO 1 Starting and Ending Points based on Annual Scale Score Increase

Grade Band	Percent Gain by Initial Proficiency Level				
	Criteria	1-2	3	4	5
K-2	Start	28	18	12	2
	End	42	25	21	20
3-5	Start	21	12	3	2
	End	33	20	14	12
6-8	Start	18	8	3	2
	End	25	14	8	6
9-12	Start	2	2	2	2
	End	20	11	9	8

AMAO 2—Findings and Recommendations

Among the many decisions required of establishing AMAO 2 decisions, two are critical: What is English proficient? and What cohorts of students should be considered when calculating the percent of students who are English proficient in districts?

We suggest the following activities be considered when setting English proficient expectations:

1. Empanel relevant stakeholder experts to define what AMAO 2 English language proficiency means,
2. Correlate the state's reading and mathematics test data to ACCESS scores and determine which performance is predictive of success on the reading and math assessments,
3. Identify ELL students who successfully participate in class without substantial English language support and identify which ACCESS proficiency level(s) that represents.

We identify the following groupings of students as possible AMAO 2 cohorts.

1. ELLs who were at a WIDA proficiency level of 2.5 or higher in the previous year,
2. ELLs who were at a WIDA proficiency level of less than 2.5 and who first enrolled in US schools four or more years ago,
3. ELLs who were at a WIDA proficiency level of less than 2.5 in the prior year and in US schools *less than* four years who reach the English proficient level in the current year.

Alternative Approaches for Determining Annual Measurable Achievement Objective

Two alternate approaches to establishing AMAO 1 expectations might also be considered. They differ from methods forwarded by Linquanti and George, and it is unclear how consistent they would be with current federal law. Certainly, they speak to the intent of Title III legislation, if not exactly to the letter.

Growth Normalization

The fist approach is called Growth Normalization. The mantra for student second language growth outlined in this paper is *lower is faster, higher is slower*. As a result, different expectations were applied to different grade-bands and proficiency levels. Another approach would be to create a common metric that "normalizes" or places different growth expectations upon the same scale. The equation below outlines how this might be done.

$$g_i = \frac{(y_{1ilk} - y_{0ilk}) - (\bar{y}_{1lk} - \bar{y}_{0lk})}{SD_{(\bar{y}_{1lk} - \bar{y}_{0lk})}}$$

Where l = initial proficiency level, e.g., <2, 2-3.5 >3.5,

k = grade level cluster, K-2, 3-5, 6-8, and 9-12,

y_{1ilk} = student i's final score at initial proficiency level l and cluster k

y_{0ilk} = student i's initial score at initial proficiency level l and cluster k.

g_i then becomes the normalized growth of student i at initial proficiency level *l* in grade cluster *k*. Essentially g_i is a z-score transformation of students' annual growth. These scores can be averaged across students and/or schools in a district and used as a metric for examining AMAO 1 performances. To be used for accountability, $(\bar{y}_{1lk} - \bar{y}_{0lk})$ and $SD_{(\bar{y}_{1lk}-\bar{y}_{0lk})}$ would be fixed at year 0 and subsequent years' normalization scores would use these fixed values. Year 0's fixed values would be the reference point for improvement. This procedure was applied with the three state sample and provided promising results. The benefits of this procedure is there would be one specific growth expectation score. That score would reflect expected gain across grades and proficiency levels.

In setting a starting point, one might choose 1 or 1 ½ standard deviations below the initial mean, and the ending point might be 1 or 1 ½ standard deviations above the initial mean. A drawback of this approach is that we are normalizing growth and loosing information about student progress as a result. We would be comparing student progress relative to the past, which may or may not reflect how students, schools or districts now grow. Nonetheless, the procedure is promising. Calculating growth normalization scores would not be overly taxing. g_i could be transformed into a scale that is meaningful to educators, and it would be relatively easy to communicate district growth expectations, which could be applied to all grades and levels.

Predictive Growth

Another method of identifying AMAO 1 expectations is to establish growth expectations by fitting a mixed linear model to available test data. This provides estimates of the differing growth trajectories for different proficiency levels in a grade band. In essence, you are predicting the nature of student growth. A necessary requirement for this model is an interval, vertically scaled scoring metric. This procedure would not apply to proficiency levels or proficiency level decimals scores. The following model outlines how one might estimate growth projections.

Level 1

$$y_{ti} = \pi_{0i} + \pi_{1i}(\text{Time})_{ti} + r_{ti}$$

Level 2

$$\pi_{0i} = \beta_{00} + \beta_{01}(\text{Prof. Level})_i + e_{0i}$$

$$\pi_{1i} = \beta_{10} + \beta_{11}(\text{Prof. Level})_i + e_{1i}$$

Combined Model

$$y_{ti} = \beta_{00} + \beta_{01}(\text{Prof. Level})_i + \beta_{10}(\text{Time})_{ti} + \beta_{11}(\text{Prof. Level})_i(\text{Time})_{ti} + e_{0i} + e_{1i}(\text{Time})_{ti} + r_{ti}$$

Where in the combined model y_{ti} represents student i's predicted ACCESS composite score; β_{00} is the predicted initial composite score score; β_{01} is the difference in starting score based on proficiency level; β_{10} is the annual gain in composite score; β_{11} is the annual gain in composite score based on proficiency level; and e_{0i}, $e_{1i}(\text{Time})_{ti}$, and r_{ti} are variance estimates associated with the model. For those interested, the figure below displays SAS code for this type of analysis.

Figure 20: SAS Code for Mixed Model Analyses

```
proc mixed data = <dataset>;
class student_id;
model scale_score = time level_c time*level_c / solution ddfm=bw
      notest;
random intercept time / subject = id type=un;
```

```
run;
```

Once estimated, predicted gains by grade band and proficiency level could be obtained. To set starting an ending points, standard errors associated with growth estimations could be applied. For example, the starting point could be set at 2 standard errors below the predicted growth estimate and the ending point could be 2 standard error values above.

These models are somewhat involved and require a degree of familiarity with mixed model techniques. It is important to note that the model shown above will be biased if corrections for students who exit early are not controlled for. As mentioned earlier, students who exit ELL programs are typically excused from taking ACCESS. Estimating student growth by proficiency level without accounting for these students will underestimate students' growth trajectories. Thus, a selection bias correction, e.g., a Heckman model correction (Heckman, 1979), will be necessary to correctly estimate student growth. Preliminary analyses using a predictive growth model were conducted using the three state data set. Results suggest that this procedure has promise; however, more research is necessary to justify the use of this procedure. Value-added modeling techniques could also be applied using this procedure. For those interested in more detail about value-added models, see the Spring 2004 edition of the *Journal of Educational and Behavioral Statistics*.

References

Cizek, G.J. (2001). Setting Performance Standards: Concepts, Methods, and Perspectives. Mahwah, NJ: Lawrence Earlbaum Associates, Publishers.

Collier, V. (1988). The Effect of Age on Acquisition of a Second Language for School. New Focus: *The National Clearinghouse for Bilingual Education, Occasional Papers in Bilingual Education, No.2*. Retrieved September 5, 2007 from www.ncela.gwu.edu/pubs/classics/focus/02aage.htm.

Collier, V. (1995). Acquiring a Second Language for School. Directions in Language & Education, National Clearinghouse for Bilingual Education, Vol.1, No. 4. Retrieved September 5, 2007 from www.ncbe.gwu.ncbepubs/directions/04.htm.

Cummins, J. (1983). Language Proficiency and Academic Achievement. In J. Oller (ed.). 1983. *Issues in Language Testing Research*. Rowley, MA.: Newbury House.

Doughty, C.J. and Long, M.H. (2003). The Scope of Inquiry and Goals of SLA. In C.J. Doughty & M.H. Long (Ed.) *The Handbook of Second Language Acquisition*. Malden, MA: Blackwell Publishing Ltd.

Gottlieb, M. (2006). *Assessing English Language Learners: Bridges from Language Proficiency to Academic Achievement.* Thousand Oaks, CA.: Corwin Press.

Hambleton, R.K. and Pitoniak, M.J. (2006). Setting Performance Standards. In R. L. Brennan (Ed.) Educational Measurement, Fourth Edition. Westport, CT: Praeger Publishers.

Hakuta, K., Butler, Y.G., and Witt, D. (2000). How Long Does It Take English Learners to Attain Proficiency? The University of California Linguistic Minority Research

Institute Technical Report 2000-1. Retrieved September 5, 2007 from http://www.lmri.ucsb.edu/publications/00_hakuta.pdf.

Heckman, James J. 1979. Sample Selection Bias as a Specification Error. Econometrica 47 (1):153-161.

Hyltenstam, K. and Abrahamsson, N. (2003). Maturational Constraints in SLA. In C.J. Doughty & M.H. Long (Ed.) *The Handbook of Second Language Acquisition*. Malden, MA: Blackwell Publishing Ltd.

Gottlieb, M. (2006). *Assessing English Language Learners: Bridges from Language Proficiency to Academic Achievement.* Thousand Oaks, CA.: Corwin Press.

Krashen, S. (1987). *Principles and Practice in Second Language Acquisition.* New York, NY: Prentice-Hall International.

Linquanti, R. & George, C. (in press). Establishing and utilizing an NCLB Title III accountability system: California's approach and findings to date. In J. Abedi (Ed.), English language proficiency assessment in the United States: Current status and future practice. Los Angeles and Davis: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST). Forthcoming at: [http://www.cse.ucla.edu/products/reports.asp].

Long, M.H. (2003). Stabilization and Fossilization in Interlanguage. In C.J. Doughty & M.H. Long (Ed.) *The Handbook of Second Language Acquisition*. Malden, MA: Blackwell Publishing Ltd.

McLaughlin, B. (1984). *Second-Language Acquisition in Childhood: Volume 2. School-Age Children Second Edition*. Hillsdale, NJ: Lawrence Earlbaum Associates, Publishers.

Mohan, B.A. (1979). Language Teaching and Content Teaching. *TESOL Quarterly* 13(2): 171 – 182.

Mohan. B.A. 1986. *Language and Content.* Reading, MA.: Addison-Wesley Publishing Co.

The No Child Left Behind Act of 2001 (NCLB). Pub. L. 107-110, Jan 8, 2002. Stat.115.1425-2094.

Scarcella, R. (2003). *Academic English: A conceptual framework*. The University of California Linguistic Minority Research Institute Technical Report 2003-1. Retrieved September 5, 2007 from http://www.lmri.ucsb.edu/publications/03_scarcella.pdf.

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics in Language Teaching, 10*: 3, 209-231.

Appendix A: Sample of WIDA State AMAO Policies