WCER Working Paper No. 2008-2
April 2008

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

H. Gary Cook
Value-Added Research Center/
Wisconsin Center for Education Research
University of Wisconsin–Madison
hcook@wisc.edu

Carsten Wilmes
WIDA Consortium/
Wisconsin Center for Education Research
University of Wisconsin–Madison
wilmes@wisc.edu

Tim Boals
WIDA Consortium/
Wisconsin Center for Education Research
University of Wisconsin–Madison
tjboals@wisc.edu

Martín Santos
WIDA Consortium/
Wisconsin Center for Education Research
University of Wisconsin–Madison
msantos@ssc.wisc.edu

WCER
Wisconsin Center for Education Research
School of Education • University of Wisconsin–Madison • http://www.wcer.wisc.edu/

WCER working papers are available on the Internet at http://www.wcer.wisc.edu/publications/workingPapers/index.php. Recommended citation:

> Cook, H. G., Boals, T., Wilmes, C., & Santos, M. (2008). *Issues in the development of annual measurable achievement objectives for WIDA consortium states* (WCER Working Paper No. 2008-2). Madison: University of Wisconsin–Madison, Wisconsin Center for Education Research. Retrieved [e.g., May 15, 2008,] from http://www.wcer.wisc.edu/publications/workingPapers/papers.php

The research reported in this paper was supported by the WIDA Consortium and by the Wisconsin Center for Education Research, School of Education, University of Wisconsin–Madison. Any opinions, findings, or conclusions expressed in this paper are those of the authors and do not necessarily reflect the views of the funding agencies, WCER, or cooperating institutions.

Issues in the Development of Annual Measurable Achievement Objectives for WIDA Consortium States[1]

H. Gary Cook, Tim Boals, Carsten Wilmes, and Martín Santos

Purpose and Background

Title III of the No Child Left Behind Act of 2001 (NCLB; 2002) 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, and writing. In states of the World-class Instructional Design and Assessment (WIDA) Consortium,[2] the NCLB-approved English language proficiency assessment is ACCESS for ELLs® (Assessing Comprehension and Communication in English State to State). This assessment is aligned to WIDA English proficiency standards and reflects the state-specific academic content standards of member states as required under NCLB.

As a consortium dedicated to assisting member states with both achieving legal compliance and improving the education of ELLs, WIDA provides guidance and support to states and local schools. The goal of this paper is to assist states as they formulate AMAOs—specifically, (a) targeted annual increases in the number or percentage of children making progress in learning English (*AMAO 1*) and (b) targeted annual increases in the number or percentage of children attaining English proficiency (*AMAO 2*). This paper will not address issues associated with AMAO 3, which primarily concerns adequate yearly progress (AYP) for ELLs.

WIDA has previously issued two papers on this 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." 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 from three that tested in spring 2005. The recommendations were tentative and based largely on national studies estimating the 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 2 years of available data from the same three states. This report was

[1] 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 and insight. Any mistakes, errors, or omissions are those of the authors.

[2] WIDA is dedicated to the design and implementation of high standards and equitable educational opportunities for English language learners. Originally established through a federal grant, the WIDA Consortium consists of 17 partner states: Alabama, Delaware, the District of Columbia, Georgia, Illinois, Kentucky, Maine, New Hampshire, New Jersey, North Dakota, Pennsylvania, Oklahoma, Rhode Island, South Dakota, Vermont, Virginia, and Wisconsin. These 17 states account for over 550,000 ELLs in kindergarten through Grade 12 in more than 3,500 school districts.

issued August 30, 2006, as "Composite Score Growth Profile Based on Years 1 & 2 ACCESS for ELLs® Data, WIDA Consortium Draft Recommendations." The analysis in this second report represented a significant refinement of the general recommendations in the first report.

WIDA also issued a "Bridge Study Report" by Dorry Kenyon in August 2006 (Technical Report #2).[3] This report compared four prior English language proficiency assessments—IDEA Proficiency Test (IPT);[4] Language Assessment System (LAS);[5] Maculaitis Assessment of Competencies II Test of English Language Proficiency (MAC II);[6] and Woodcock-Muñoz Language Survey[7]—to scores on ACCESS for ELLs®. The Bridge data were used by at least three WIDA states in setting their initial AMAOs for ACCESS for ELLs®.

WIDA now has 3 years of results from the original three states and 2 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 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* had emerged, albeit with less empirical support than it has today (Cummins, 1983). In 1979, Mohan's article "Relating Language Teaching and Content Teaching" was published in *TESOL Quarterly*. Mohan's seminal text *Language and Content* was published in 1986, and the modern movement to teach language through content had begun.

The heightened awareness of the unique demands of academic English coincided with the recognition that full English language proficiency was likely to take much longer to attain than originally posited. Students who were able to carry on good conversations in English within a year or two of first enrolling in U.S. schools struggled for a substantially longer time (up to 7 years) to learn to use English effectively in academic classes. Some students with weaker literacy skills in their primary language or with limited school experience appeared to require even more time to close the academic content and language gaps. The recognition of these lags in mastering academic English led, in turn, to a recognition of the need to teach language through content, since English language learners could not afford to spend large chunks of the day mastering English while their English proficient peers continued to progress in mathematics, science, and

[3] This report is available in the technical reports section of the ACCESS for ELLs® Web page on the consortium Web site (www.wida.us).

[4] www.ballard-tighe.com/assessment.html

[5] www.ctb.com/products/category_home.jsp?FOLDER%3C%3Efolder_id=2534374302057165&bmUID%20=1156189729545

[6] store.cambiumlearning.com/ProgramPage.aspx?parentId=019005529&functionID=009000008&pID=&site=sw

[7] www.assess.nelson.com/test-ind/wmls-r.html

other subjects. By the time NCLB was enacted in 2001, the force of federal law backed the idea that ELLs must be included in school-level accountability and hence assessed in both academic content and English proficiency. Gone were the days when schools had the luxury of teaching English to ELLs and worrying about helping them catch up with everything else later.

Prior to NCLB, few practical tools were available to assist teachers in making the transition to teaching grade-appropriate content along with subject-specific English. NCLB's requirement that states develop English proficiency standards aligned with academic language and linked to academic content standards has provided a blueprint for materials development, lesson planning, classroom assessment, and staff development. The English proficiency standards also serve as the blueprint for English proficiency tests. Ideally, all standards and assessments within the system push schools in the same direction, toward including ELLs in grade-appropriate curriculum designed to maximize their learning of content while increasing their proficiency in English. This is what WIDA's English language proficiency standards seek to accomplish.

To assess WIDA standards, the consortium developed ACCESS for ELLs® (ACCESS). ACCESS is composed of four domain-based tests (speaking, listening, reading, and writing) that assess students at four grade-level clusters: (a) Grades K–2, (b) Grades 3–5, (c) Grades 6–8, and (d) Grades 9–12. For each domain and grade-level cluster (except K), ACCESS is divided into three tiers. Tier A assessments target students at the three lowest language proficiency levels (Levels 1–3); Tier B, students at the middle proficiency levels (Levels 2–5); and Tier C, students at the higher proficiency levels (Levels 3–6). Psychometrically, ACCESS is vertically scaled across grades and proficiency levels using a Rasch-based item response theory (IRT) scaling methodology. The ACCESS scale ranges from 100 to 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 academic language proficiency as specified in the WIDA standards. For more information on the reliability and validity of ACCESS, see www.wida.us/assessment/ACCESS_techReports/index.aspx.

Data Used for Analyses

The data used for this report come from three WIDA Consortium states over a 3-year period (2005–2007) across all grades (kindergarten to 12th grade). A total of 12,836 ELL students are included in the analyses: 9,542 from State A, 2,154 from State B, and 1,140 from State C. In 2005, all 12,836 students were included in the sample. A total of 12,014 students remained in the 2006 sample, and 9,353, in the 2007 sample. The reduction in student numbers over time resulted from missing data or from students' exiting ELL programs, moving, or missing the assessment. It is important to note that this is a longitudinal data set: students included in this sample have at least two data points, and more than 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, as shown in Table 1.

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%

Note. From *ACCESS for ELLs®: Interpretive Guide for Score Reports*, by the WIDA Consortium, 2007, p. 8 (www.wida.us/assessment/ACCESS%20Interpretive%20Guide07.pdf). Copyright 2007 by the Board of Regents of the University of Wisconsin System, on behalf of the WIDA Consortium. Reprinted with permission.

Rasch model–generated scale scores are created for the four domain tests. Domain test scale scores 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 score. The composite score is a domain-based weighted score, interpreted from scale scores and derived in the following manner: speaking (15%), listening (15%), reading (35%), and writing (35%). Thus, 70% of the composite score reflects text-based proficiency, and 30%, oral/aural proficiency. All AMAO analyses reported in this paper use the composite score since most WIDA states use this as the metric for AMAO expectations (see Appendix A).

Before moving to a discussion of setting AMAO targets, we first briefly review what we know about ELL students' growth in English.

Research on Second Language Acquisition

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*. Demanding that language programs "move students along" in some cases belies the very nature of language learning and is unreasonable and unrealistic. The converse is also problematic—when student growth expectations are set so low that students are permitted to take inordinately long to learn English. According to Scarcella (2003), such low expectations have severe economic consequences 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.

Research[8] suggests that students' second language acquisition is influenced by a variety of factors, such as social context, language aptitude, types of inputs and interactions, cognitive processing, first language, and prior educational experiences. Two factors are of particular

[8] It is beyond the scope of this paper to overview child second language research in detail. For more information about child and adult second language acquisition, see McLaughlin (1984) and Doughty and Long (2003).

interest in setting AMAOs: (a) the maturational constraints on learning a second language (see Hyltenstam & Abrahamsson, 2003; Collier, 1987–1988, 1995); and (b) the stabilization of language development (Long, 2003).

It has been 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 is so. One popular hypothesis is that there is a critical period (see Birdsong, 1999), typically before puberty, in which language learning can occur with little effort. However, the critical period hypothesis is by no means universally accepted. Many have argued that adults can acquire language fluently, but at dramatically different rates than children. Still others have suggested that learning environments can play a significant role in facilitating acquisition—for example, 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 to be seen in older children—make a significant difference in the rate of second language acquisition (Collier, 1995). These caveats aside, most agree that younger learners learn at higher rates than older learners.

When considering the stabilization of language development, researchers observe that as language learners move to higher levels of proficiency, the rate at which language is acquired slows down. In many instances, language learning can "fossilize" or, as Long (2003) puts it, "stabilize." That is, language learners operate at an interlanguage (Selinker, 1972) level. *Interlanguage* refers to something between the first and second language. It is not the fluent use of the second language, nor is it confined to first language structures; it is something in between. The process of 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 about whether children stabilize in their language development and, if so, at what age. 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—maturational constraints and stabilization of language development—can be combined into the following principle: *lower is faster, higher is slower*. That is, language learners in lower grades (younger students) and at lower proficiency levels acquire language at faster rates. Thus, if two students were at the same proficiency level but in different grades, we would predict that the younger student would acquire language at a faster rate than the older student. This principle should inform the setting of AMAO expectations.

Development of AMAO 1: Progress in Learning English

A premise behind AMAO 1 (the number or percentage of ELLs making progress in learning English) is that districts charged with teaching English to ELLs must show continuous improvement. There are two elements to this improvement: first, students must make appropriate progress in their language learning, and second, district ELL programs must continue to improve the rate at which students make progress. In this section, we use data from three WIDA states to outline a process for establishing AMAO 1 goals. All three states' ACCESS composite scores are combined to form one data set. The methods we outline are intended to serve consortium members as guides as they set their own AMAO 1 goals.

The model we recommend and use here for setting AMAO targets is that set forth by Linquanti and George (2007). Linquanti and George described 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). Table 2 identifies the federal statutes underlying Linquanti and George's procedures.

Table 2
AMAO: NCLB Definitions

	Definition	Citation
AMAO	Each State educational agency or specially qualified agency receiving a grant under subpart 1 of this part 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 [20 U.S.C. § 6311(b)(1)].	20 U.S.C. § 6842(a)(1)
AMAO categories	Such annual measurable achievement objectives . . . shall include . . .	
AMAO 1—Progress	at a minimum, annual increases in the number or percentage of children making progress in learning English;	20 U.S.C. § 6842(a)(3)(A)(i)
AMAO 2—Proficiency	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 [20 U.S.C. § 6311(b)(7)]; and	20 U.S.C. § 6842(a)(3)(A)(ii)
AMAO 3—AYP	making adequate yearly progress for limited English proficient children as described in [20 U.S.C. § 6311(b)(2)(B)].	20 U.S.C. § 6842(a)(3)(A)(iii)

Linquanti and George (2007, p. 106) identified five key decisions that must be made in order to establish AMAO 1 expectations. They are:

1. Determine the scoring metric to be used to measure growth.

2. Determine the annual growth target.

3. Set the starting point for AMAO 1 targets.

4. Set the ending point for AMAO 1 targets.

5. Determine the annual rate of growth.

We discuss these decisions below.

Determine Scoring Metric

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

1. WIDA English Language Proficiency Levels 1–5 (Entering, Beginning, Developing, Expanding, Bridging);

2. WIDA English language proficiency level decimal scores (1.1 to 5.9);[9] and

3. Scale scores (Rasch-based IRT vertically scaled scores from kindergarten to 12th grade ranging from 100 to 600).

Both proficiency levels and proficiency level decimal scores were established using traditional standard-setting techniques.[10] The WIDA proficiency scores used here are the most recent grade-specific cut points, which were reset in 2007. New WIDA proficiency cut scores (based on individual grades rather than grade-level clusters) are back-applied to 2005 and 2006 data. Thus, the analyses reported here provide apples-to-apples comparisons. The analyses use only ACCESS composite scores. As noted earlier, the composite score is a weighted version of the four ACCESS domain scores (speaking 15%, listening 15%, reading 35%, and writing 35%).

Since WIDA Consortium members vary in the metrics they use to determine AMAOs, we analyze all three metrics. For examples of WIDA states' AMAO policies, see Appendix A. For clarity in presentation, we discuss each metric in a separate section below.

Determine Annual Growth Target

Proficiency Level Gain

Tables 3 and 4 and Figure 1 show within-student (i.e., within-cohort) growth across the 2005, 2006, and 2007 school years aggregated across analyzed states. States that use the proficiency level metric often base AMAO 1 expectations on the percentage of students who gain one or more proficiency levels per year. Tables 3 and 4 show the percentage of students who gained one or more WIDA proficiency levels by grade band and proficiency level.

[9] As explained in WIDA Consortium (2007), "The whole number [in the proficiency level decimal score] 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" (p. 6).

[10] For more information on this process, see the WIDA technical reports at www.wida.us/assessment/ACCESS_techReports/index.aspx.

Table 3
Percentage 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 4
Percentage 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 3 and 4 support the principle stated earlier—that is, "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 principle by aggregating the 2 years. This pattern suggests that AMAO 1 targets should not be "one size fits all," but rather should be formulated to reflect the different language acquisition rates of the different grade bands and proficiency levels.

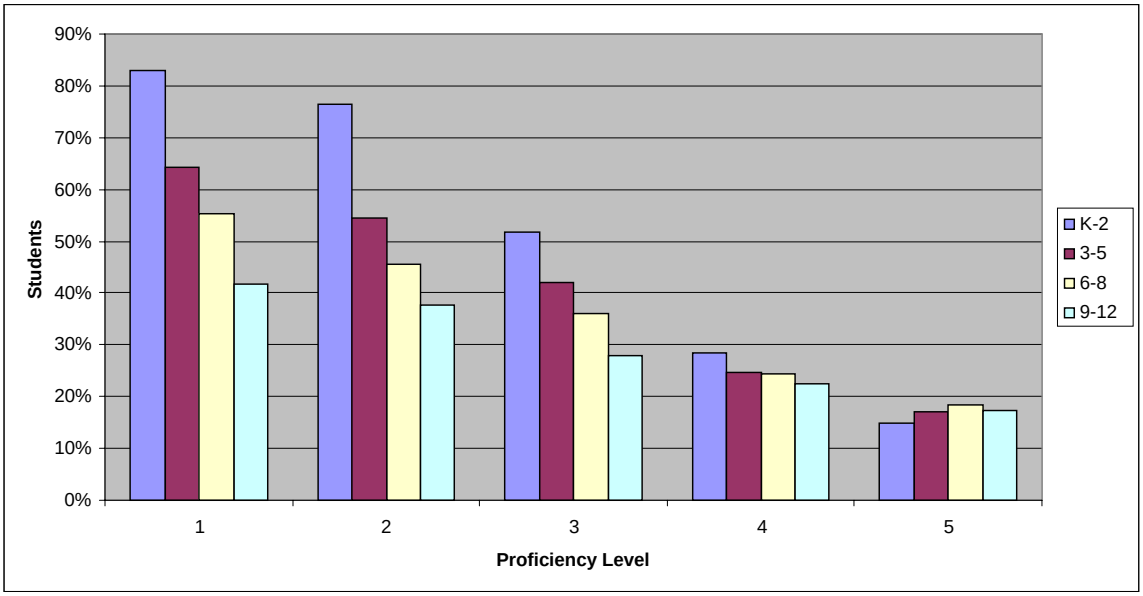
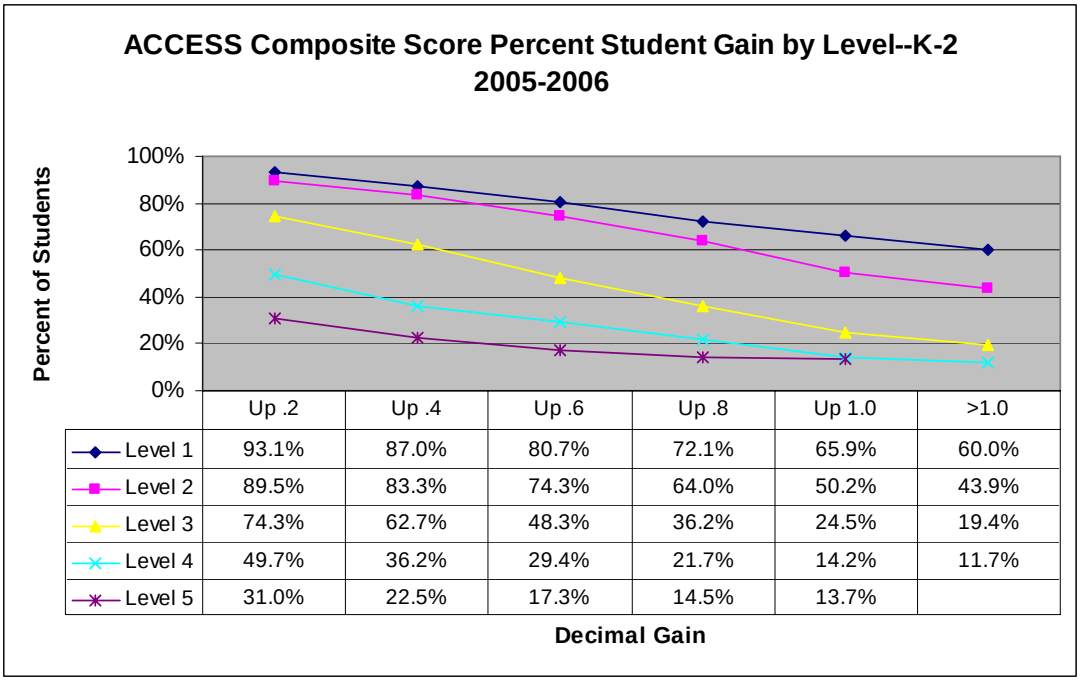


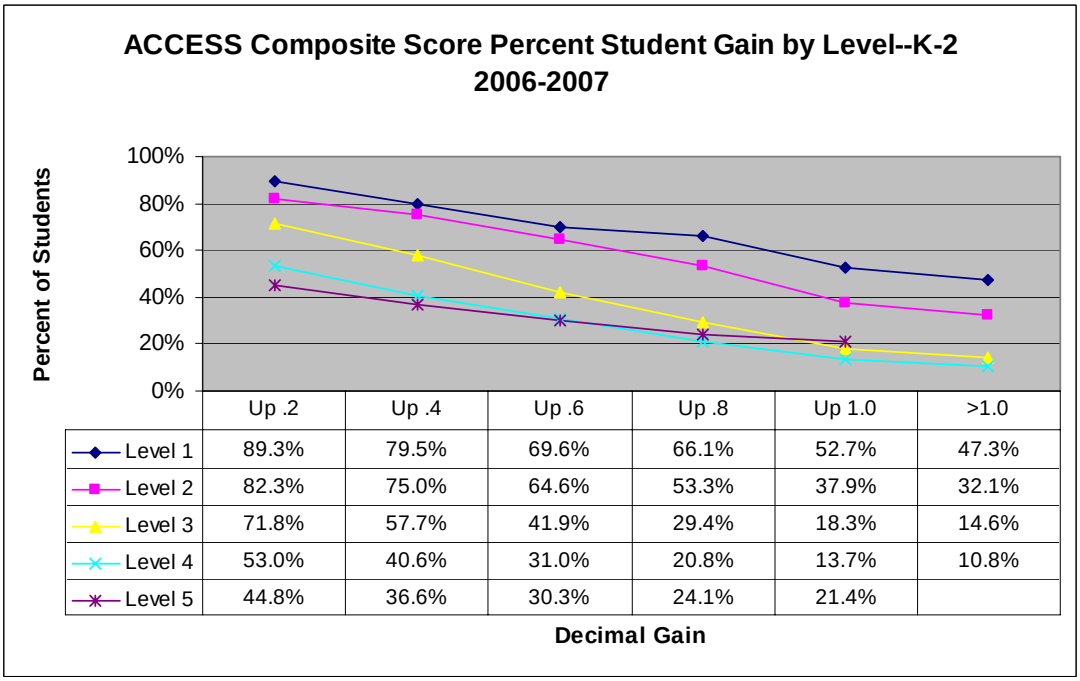
Figure 1. Aggregate percentage of students gaining one or more proficiency levels: 2005–2007.

Proficiency Level Decimal Gain

Some states set AMAO targets based on proficiency level decimal increases. Figures 2–5 show the percentage of students gaining in composite proficiency levels in 0.2 decimal increments.

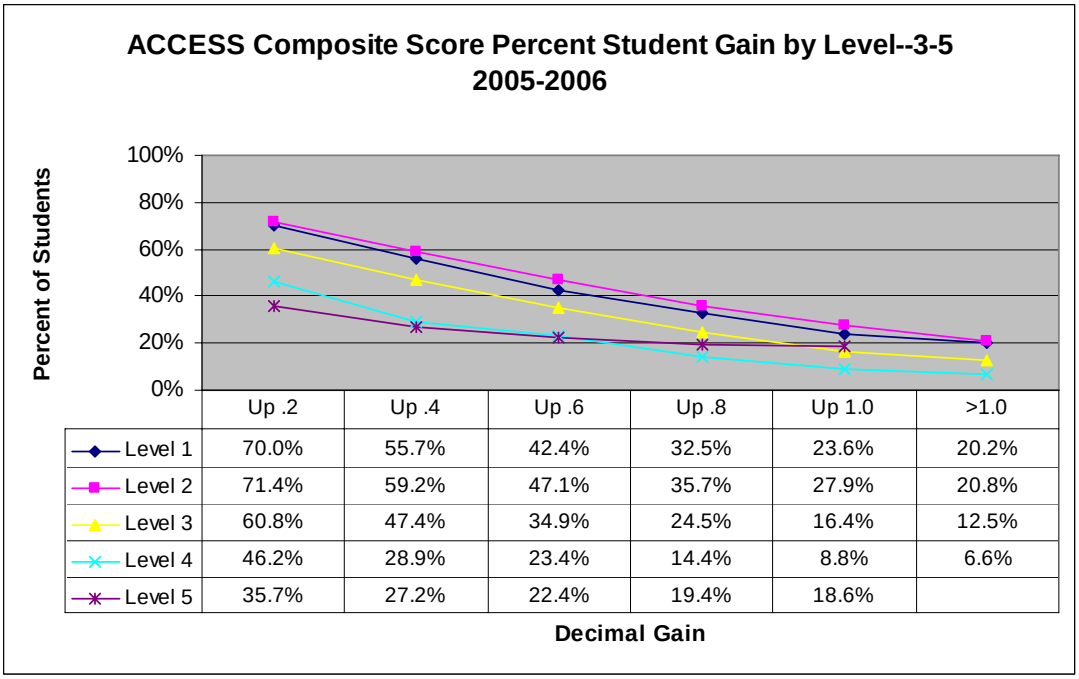


	Up .2	Up .4	Up .6	Up .8	Up 1.0	>1.0
Level 1	93.1%	87.0%	80.7%	72.1%	65.9%	60.0%
Level 2	89.5%	83.3%	74.3%	64.0%	50.2%	43.9%
Level 3	74.3%	62.7%	48.3%	36.2%	24.5%	19.4%
Level 4	49.7%	36.2%	29.4%	21.7%	14.2%	11.7%
Level 5	31.0%	22.5%	17.3%	14.5%	13.7%	

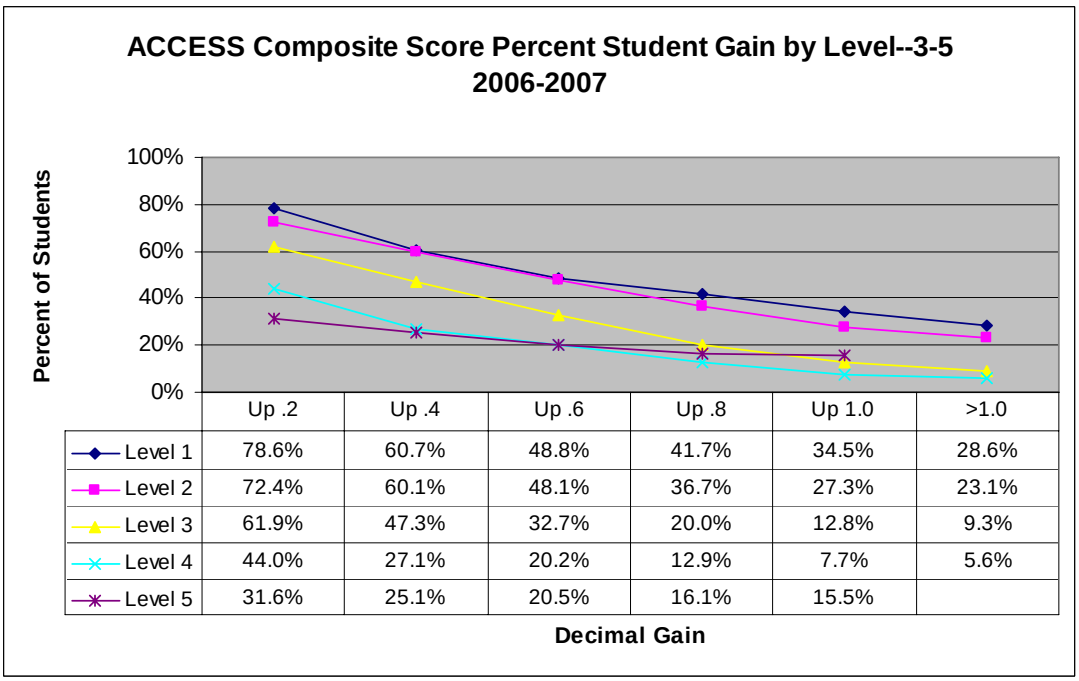


	Up .2	Up .4	Up .6	Up .8	Up 1.0	>1.0
Level 1	89.3%	79.5%	69.6%	66.1%	52.7%	47.3%
Level 2	82.3%	75.0%	64.6%	53.3%	37.9%	32.1%
Level 3	71.8%	57.7%	41.9%	29.4%	18.3%	14.6%
Level 4	53.0%	40.6%	31.0%	20.8%	13.7%	10.8%
Level 5	44.8%	36.6%	30.3%	24.1%	21.4%	

Figure 2. Composite score proficiency gain in 0.2 increments**:** K–2 grade band.

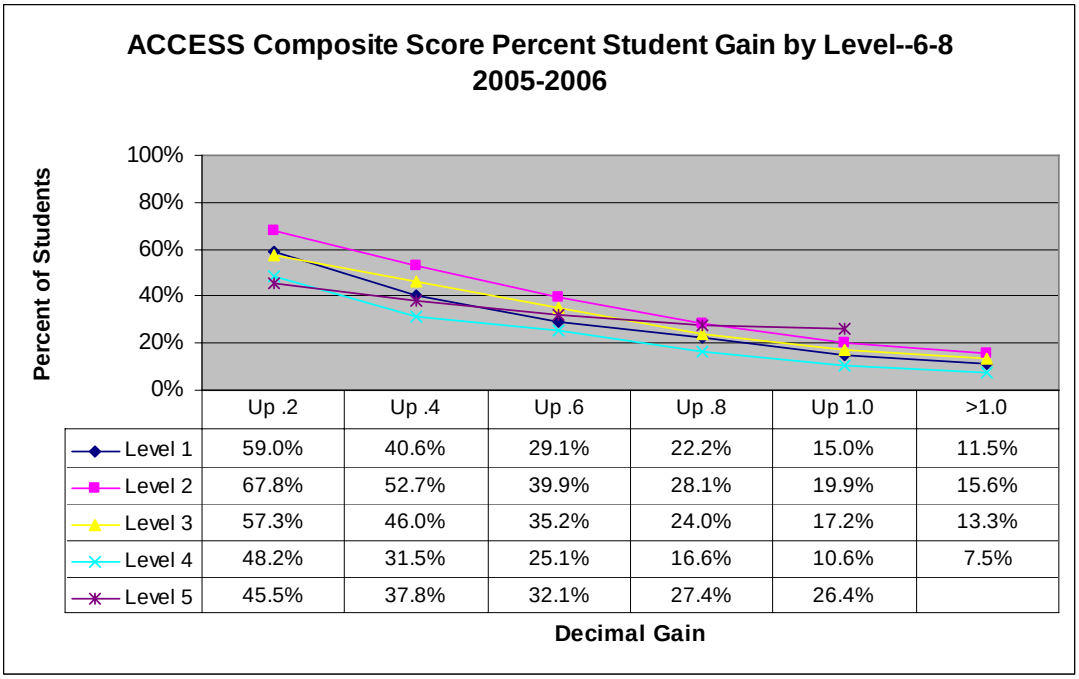


	Up .2	Up .4	Up .6	Up .8	Up 1.0	>1.0
Level 1	70.0%	55.7%	42.4%	32.5%	23.6%	20.2%
Level 2	71.4%	59.2%	47.1%	35.7%	27.9%	20.8%
Level 3	60.8%	47.4%	34.9%	24.5%	16.4%	12.5%
Level 4	46.2%	28.9%	23.4%	14.4%	8.8%	6.6%
Level 5	35.7%	27.2%	22.4%	19.4%	18.6%	

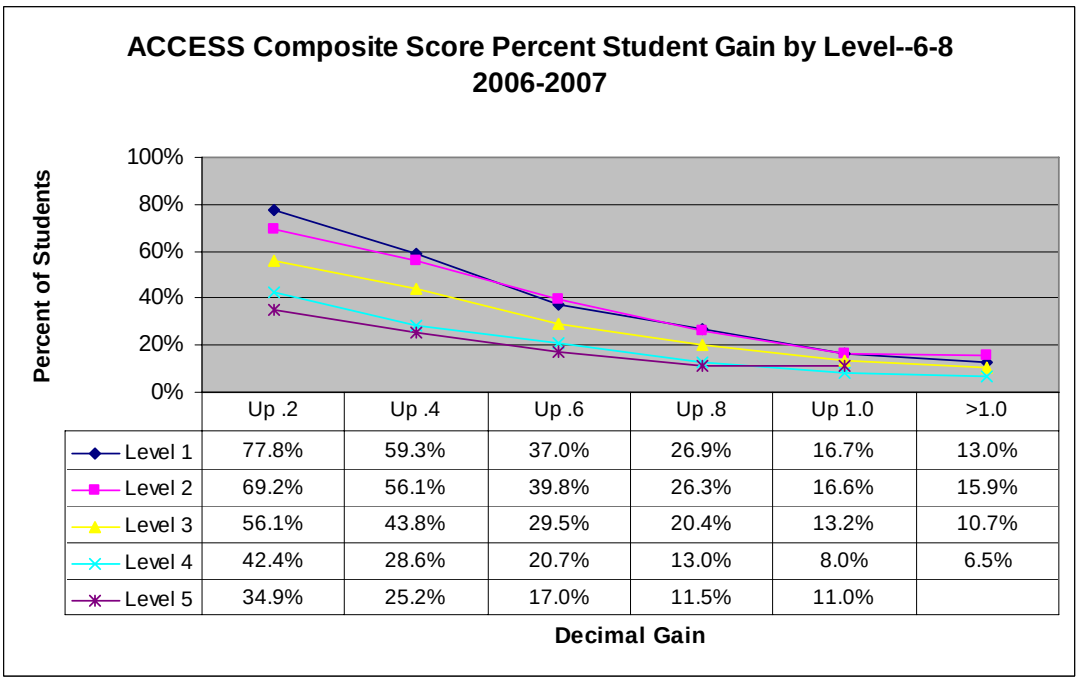


	Up .2	Up .4	Up .6	Up .8	Up 1.0	>1.0
Level 1	78.6%	60.7%	48.8%	41.7%	34.5%	28.6%
Level 2	72.4%	60.1%	48.1%	36.7%	27.3%	23.1%
Level 3	61.9%	47.3%	32.7%	20.0%	12.8%	9.3%
Level 4	44.0%	27.1%	20.2%	12.9%	7.7%	5.6%
Level 5	31.6%	25.1%	20.5%	16.1%	15.5%	

Figure 3. Composite score proficiency gain in 0.2 increments**:** 3–5 grade band.

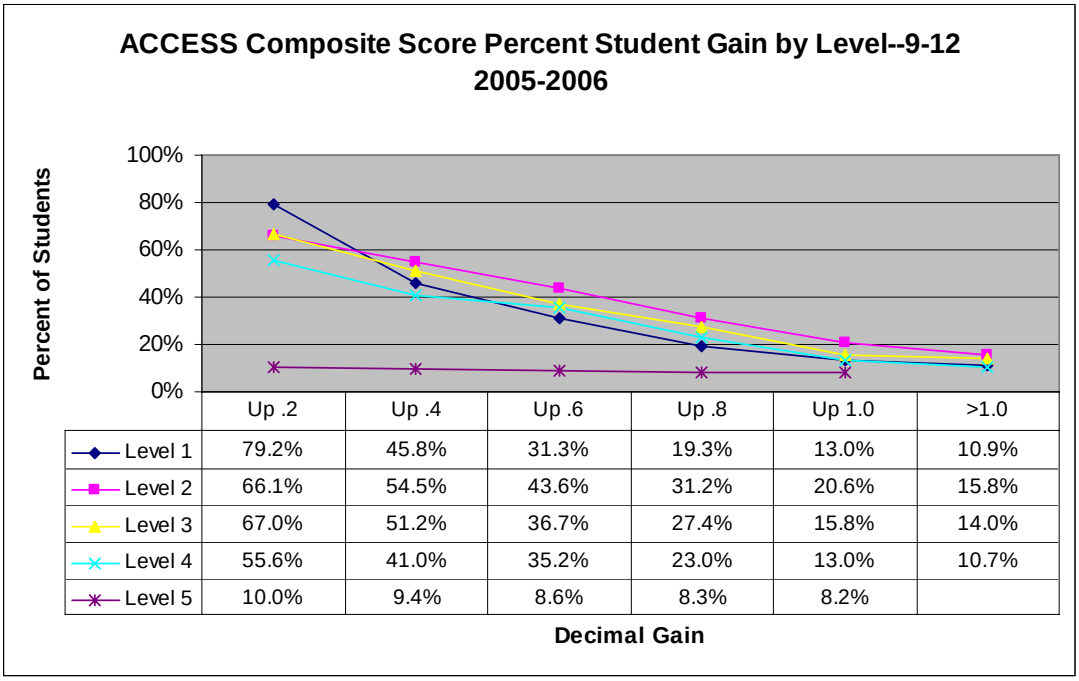


	Up .2	Up .4	Up .6	Up .8	Up 1.0	>1.0
Level 1	59.0%	40.6%	29.1%	22.2%	15.0%	11.5%
Level 2	67.8%	52.7%	39.9%	28.1%	19.9%	15.6%
Level 3	57.3%	46.0%	35.2%	24.0%	17.2%	13.3%
Level 4	48.2%	31.5%	25.1%	16.6%	10.6%	7.5%
Level 5	45.5%	37.8%	32.1%	27.4%	26.4%	

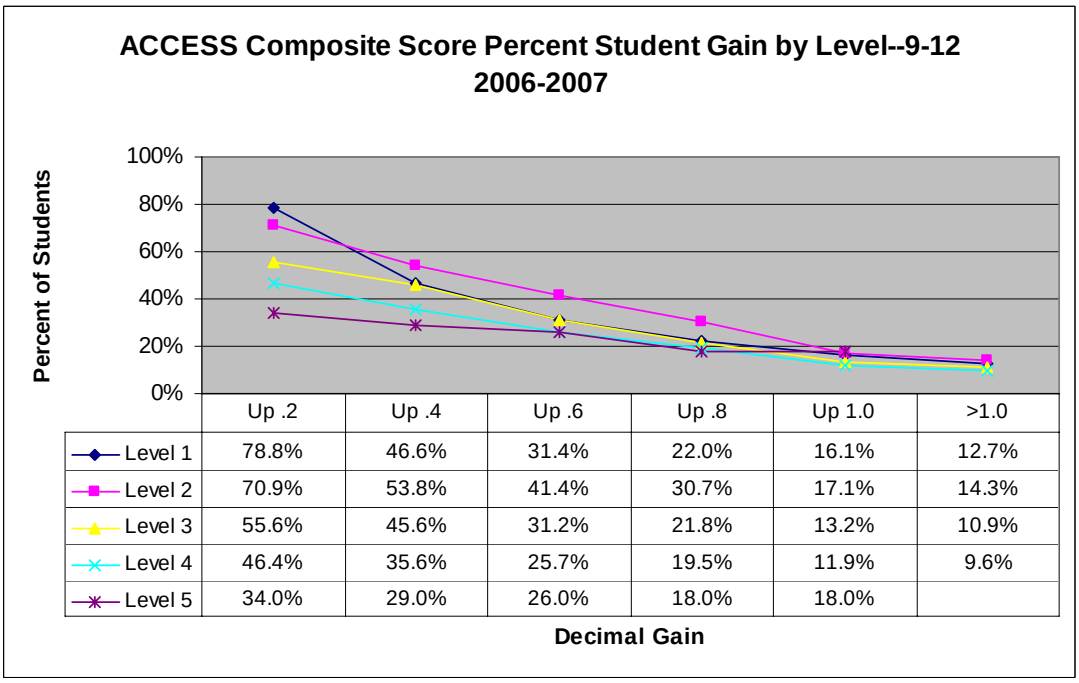


	Up .2	Up .4	Up .6	Up .8	Up 1.0	>1.0
Level 1	77.8%	59.3%	37.0%	26.9%	16.7%	13.0%
Level 2	69.2%	56.1%	39.8%	26.3%	16.6%	15.9%
Level 3	56.1%	43.8%	29.5%	20.4%	13.2%	10.7%
Level 4	42.4%	28.6%	20.7%	13.0%	8.0%	6.5%
Level 5	34.9%	25.2%	17.0%	11.5%	11.0%	

Figure 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	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%	



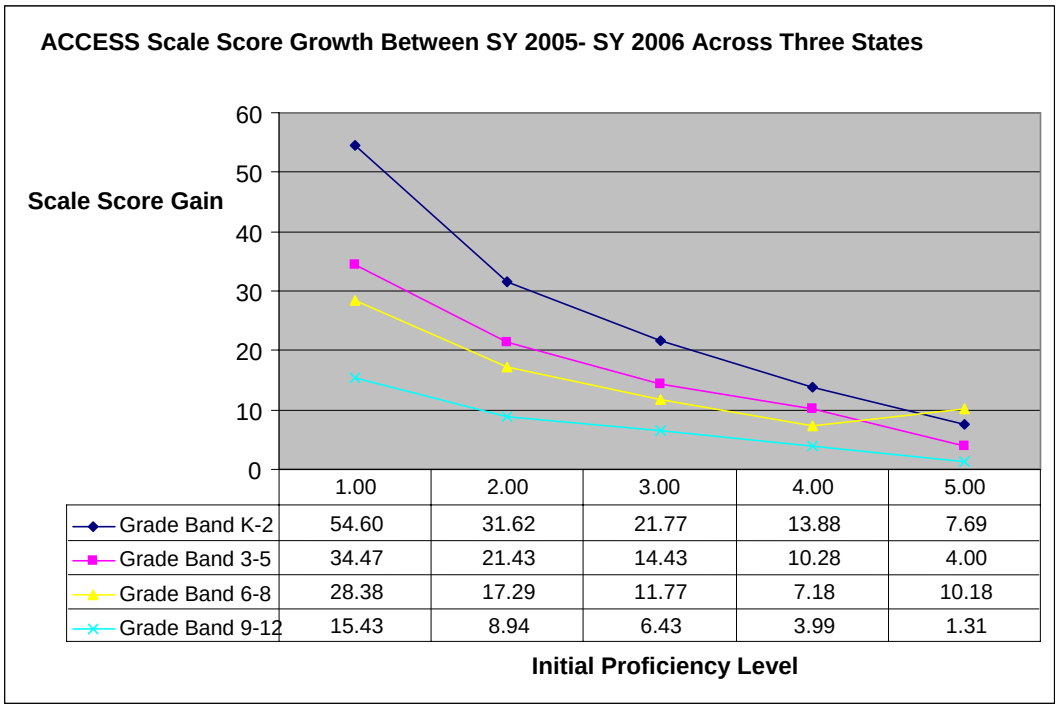
	Up .2	Up .4	Up .6	Up .8	Up 1.0	>1.0
Level 1	78.8%	46.6%	31.4%	22.0%	16.1%	12.7%
Level 2	70.9%	53.8%	41.4%	30.7%	17.1%	14.3%
Level 3	55.6%	45.6%	31.2%	21.8%	13.2%	10.9%
Level 4	46.4%	35.6%	25.7%	19.5%	11.9%	9.6%
Level 5	34.0%	29.0%	26.0%	18.0%	18.0%	

Figure 5. Composite score proficiency gain in 0.2 increments**:** 9–12 grade band.

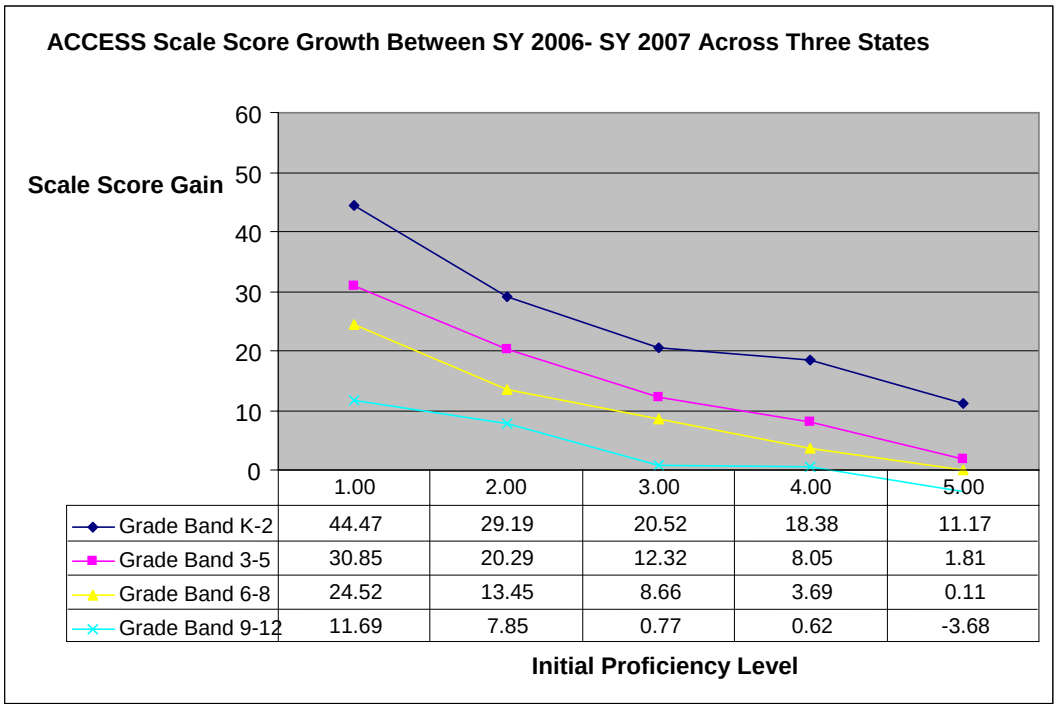
Not surprisingly, the trend for proficiency level decimal scores is similar to that for proficiency levels—that is, as the initial proficiency level increases, the percentage of students obtaining higher scores decreases. Again, these findings suggest that setting different expectations based on students' proficiency levels and grade bands may be in order.

Scale Score Gain

The graphs in Figure 6 show mean scale score gains 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 scale scores from the 2006 scale scores.



	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

Figure 6: Average scale score gain by grade band and proficiency level.

Notice in the 2006–2007 graph and table that higher proficiency levels in the 6–8 and 9–12 grade bands have very small, if not negative, scale score gains. These findings likely reflect the selectivity of the sample. Students who exit ELL programs do not take ACCESS and thus are omitted from the 2006–2007 sample; only students with ACCESS scores for all 3 years remain. It is very likely that students at higher proficiency levels who stay longer than expected in an ELL program may have motivational or special language development issues. Thus, we would expect the 2005–2006 graph to be more reflective of annual scale score gain.

The trend portrayed by these data is similar to that found for proficiency levels and proficiency level decimals—that is, students at lower grade bands and proficiency levels have greater annual scale score gains than their higher grade or proficiency level peers.

Summary

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.

These findings underscore the wisdom of tailoring annual growth targets to students' initial grade and proficiency level, regardless of the metric used.

Determine Starting and Ending Points for AMAO 1 Targets

A variety of techniques could be applied to establish starting and ending points for AMAO 1 targets. To determine the starting point for California's AMAO 1 criteria, Linquanti and George used one of the Title I methods for establishing AYP—namely, "the percentage of students at the proficient level who are in . . . 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" (20 U.S.C. § 6311(b)(2)(E)(ii)). Linquanti and George suggested three possible 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 1 targets in a similar fashion. That is, the AMAO 1 starting point is the percent/score at which the district 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 adopt the 20th and 75th percentiles.

States must also adopt minimum cell sizes for establishing AMAO criteria. The term *minimum cell size* refers to the number of students needed to make a valid and reliable decision for accountability purposes. In California's case, that number was 25—that is, at least 25 ELL students must be enrolled in a district and have two data points to be in AMAO rankings. States must adopt similar types of *n*-size criteria. In line with the growth expectations we are

recommending for the various proficiency levels and grade bands, our criterion here is 5 in each cell and 20 or more ELL students in the district. We suggest that this is a minimum criterion. We also adopt the most recent year of data (2006–2007) to rank districts.

Proficiency Level Starting and Ending Points

Table 5 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 to increase cell sizes; our analyses suggest that this is a reasonable strategy.

Table 5
Rankings of the Percentage 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 6 shows the AMAO 1 starting and ending points for proficiency levels. It may be desirable to smooth data to communicate AMAO 1 criteria more clearly and to ensure that higher grades or proficiency levels do not have more stringent criteria. For example, the percentage of students gaining one or more proficiency levels in districts at the 75th percentile for Level 4 in the 6–8 grade band is 28.1%. The percentage at the 9–12 grade band is higher—33.3%. 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.

Table 6
Starting and Ending Points for AMAO 1 Using the Percentage 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[a]	25.0[a]	15[a]	25.0[a]	25

[a]Too few districts had sufficient numbers of students at this grade and level.

Proficiency Level Decimal Starting and Ending Points

Table 7 presents percentile rankings for the average decimal gain by district for each grade band and proficiency level.

Table 7
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
K–2	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

Band	Levels	P20	P25	P50	P75	P90
	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

Table 7 shows 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 at Proficiency Level 5 from districts at the 20th percentile for the 9–12 grade band decreased by 0.73 proficiency level decimal units between the 2006 and 2007 school years. This negative growth in decimal scores may occur for a variety of reasons:

1. *Selectivity of the sample.* Students who exit from ELL programs do not take ACCESS and hence are excluded from this sample. Thus, students at higher proficiency levels who stay longer than expected in an ELL program may constitute a subpopulation facing particular problems (e.g., cognitive, family, language acquisition, primary language literacy skills).

2. *Ceiling effects.* Students at higher proficiency levels do not have much room to grow as they are approaching the ACCESS test's upper bound (Proficiency Level 6).

3. *Unique characteristics of WIDA proficiency decimal scores.* Recall that decimal 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 one between 3.2 and 4.2.

4. *District-level factors.* While students at higher proficiency levels from districts at the higher percentiles (for instance, 90th percentile) still make some progress, this is not the case for students from districts at lower percentiles (for instance, 25th percentile). This suggests the existence of district-level contextual factors affecting ELLs' performance (e.g., the quality of the instruction of higher proficiency ELLs, test motivation of students at higher proficiency levels who do not meet other criteria for reclassification).

5. *ACCESS tier caps.* In 2006, caps (maximum levels of English language proficiency) were placed on ACCESS tiers. A student taking a Tier A test (Proficiency Levels 1–3) could not exceed Proficiency Level 4 in listening and reading; likewise, a student taking a Tier B test (Proficiency Levels 2–5) could not exceed Proficiency Level 5 in those language domains.

6. *State ELL retesting policies.* Some states may have additional criteria, other than ACCESS, to determine whether students are English proficient. In these cases, students who meet the ACCESS proficiency criteria may need to retake the test to meet other proficiency criteria, potentially becoming less motivated to perform well on subsequent test administrations.

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 show some gain, even if modest, and this should also be the expectation at the district level. As noted above, if a state has multiple criteria for reclassification, students who have met the ACCESS proficiency criteria may need to retake the test to meet other proficiency criteria. In this situation, it may be prudent to allow students who have met ACCESS proficiency to stay proficient in the subsequent test administrations and to be considered to be making progress under AMAO 1.

Table 8 shows the results of smoothing using proficiency decimal scores, at the 20th and 75th percentiles, for each grade-level cluster.

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

Table 9
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
K–2	1–2	28	28	33	42	43
	3	18	18	21	25	32
	4	12	13	18	21	27

Band	Levels	P20	P25	P50	P75	P90
	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 those from decimal gains are observed—that is, 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 10 displays smoothed scale scores for starting and ending points.

Table 10
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

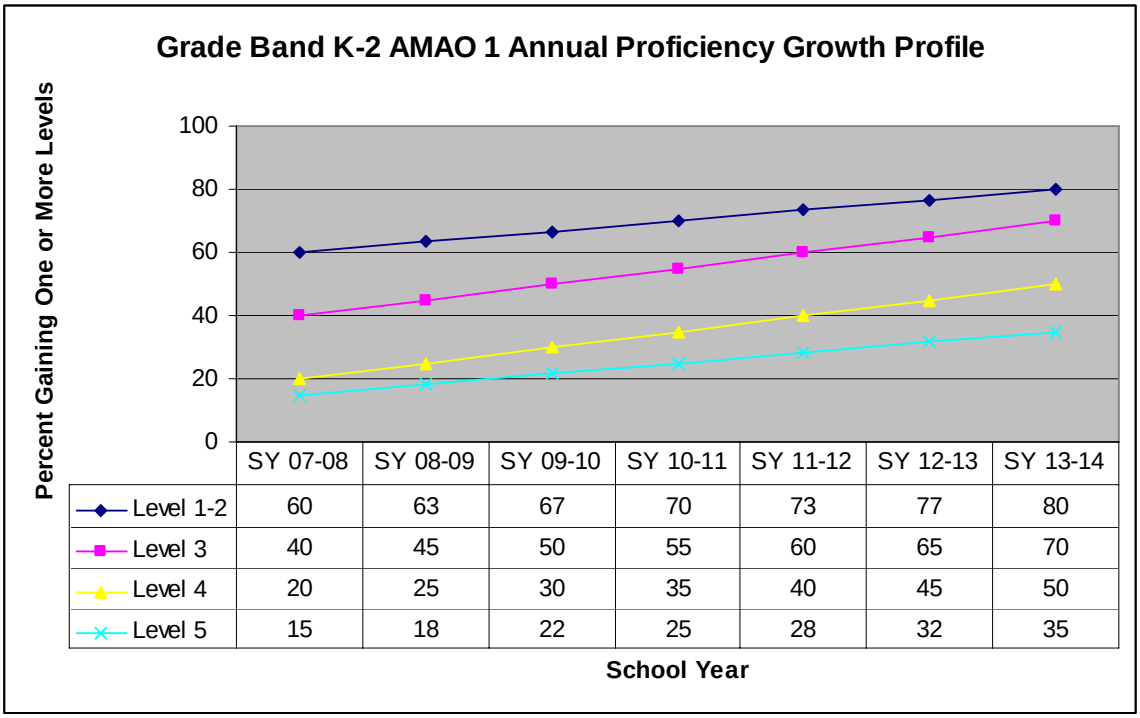
The smoothing process for scale scores focuses primarily on scores that are 1 or less. More smoothing may be desired—for example, rounding down or up to the nearest 5.

Determine Annual Rate of Growth from 2006 to 2014

Once starting and ending points are established, annual increases in district growth can be projected. Figures 7–18 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 the NCLB deadlines for AYP. States may choose to adopt other projected time frames.

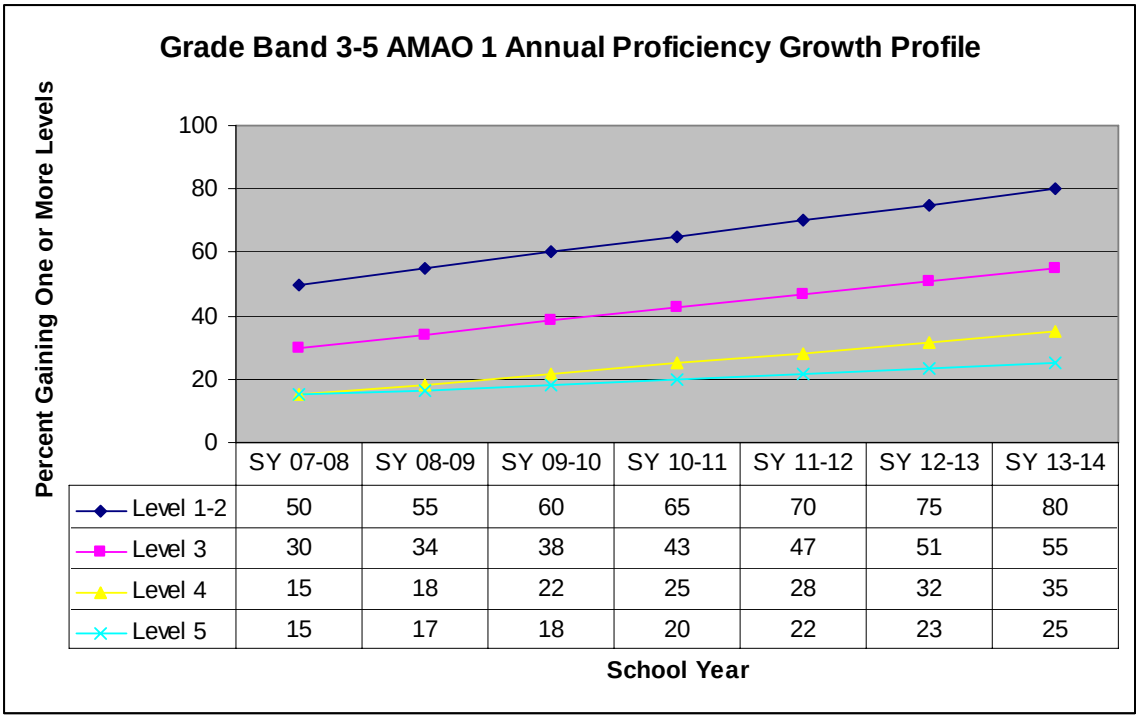
Proficiency Level AMAO 1 Profiles

Values in cells in Figures 7–10 are the percentage of students in districts, by prior level and grade band, who should gain one or more proficiency levels per year. AMAO 1 growth profiles in these 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.



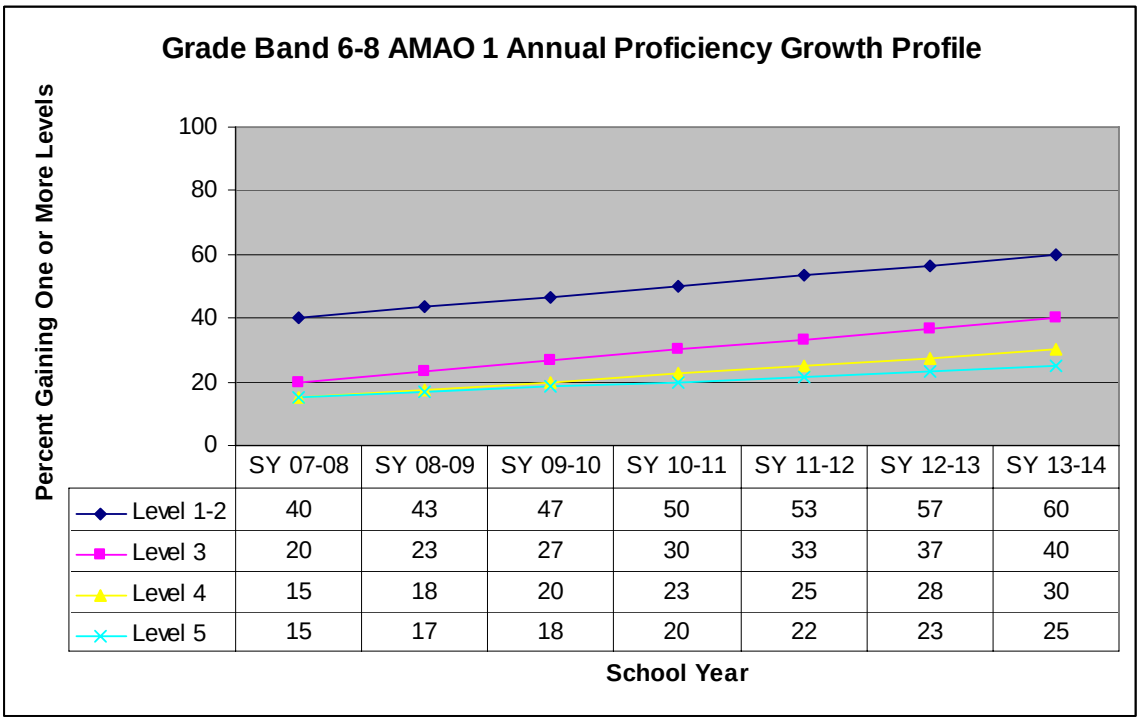
	SY 07-08	SY 08-09	SY 09-10	SY 10-11	SY 11-12	SY 12-13	SY 13-14
Level 1-2	60	63	67	70	73	77	80
Level 3	40	45	50	55	60	65	70
Level 4	20	25	30	35	40	45	50
Level 5	15	18	22	25	28	32	35

Figure 7. K–2 AMAO 1 profile for proficiency scores.



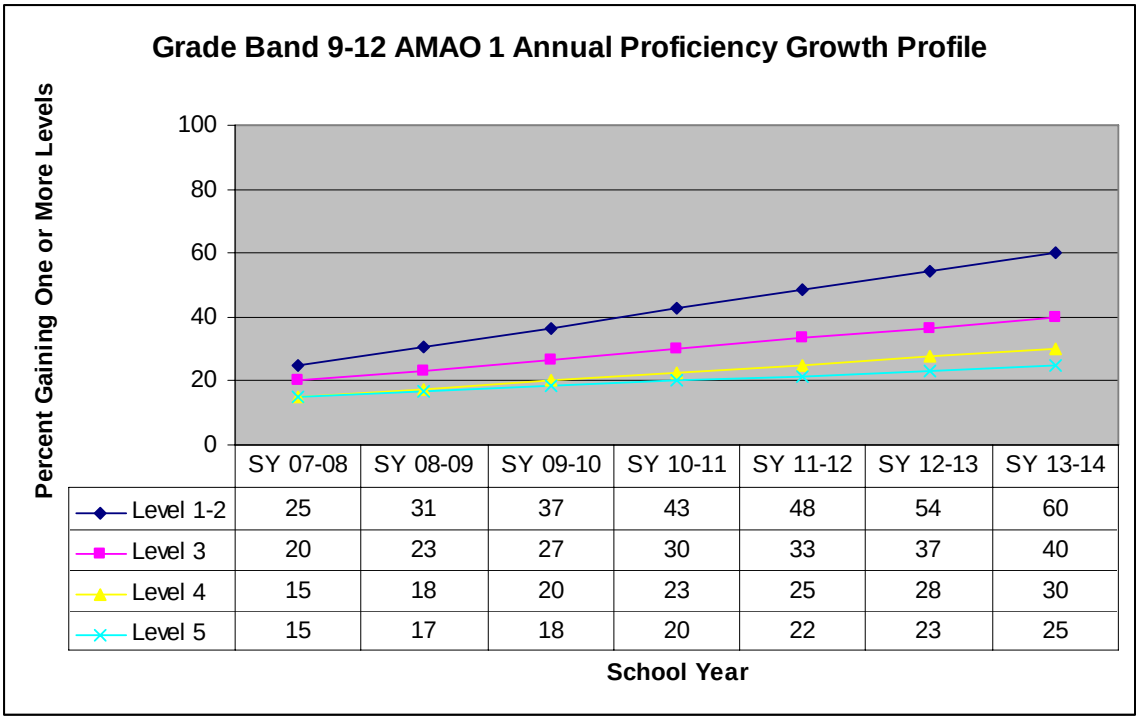
	SY 07-08	SY 08-09	SY 09-10	SY 10-11	SY 11-12	SY 12-13	SY 13-14
Level 1-2	50	55	60	65	70	75	80
Level 3	30	34	38	43	47	51	55
Level 4	15	18	22	25	28	32	35
Level 5	15	17	18	20	22	23	25

Figure 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	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 9. 6–8 AMAO 1 profile for proficiency scores.

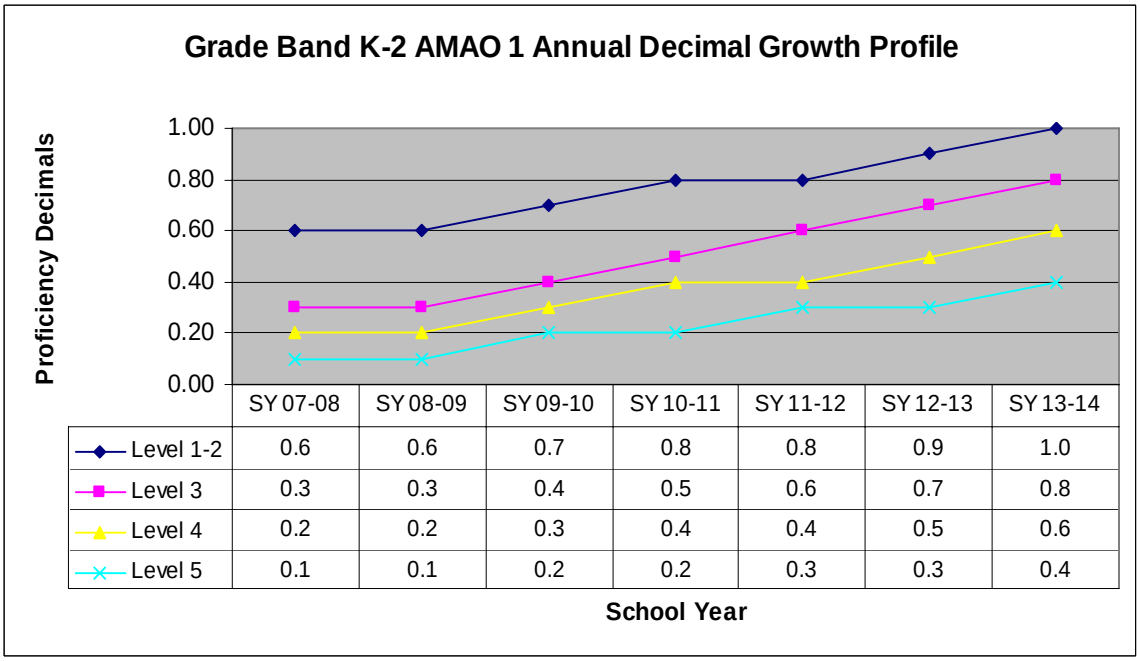


	SY 07-08	SY 08-09	SY 09-10	SY 10-11	SY 11-12	SY 12-13	SY 13-14
Level 1-2	25	31	37	43	48	54	60
Level 3	20	23	27	30	33	37	40
Level 4	15	18	20	23	25	28	30
Level 5	15	17	18	20	22	23	25

Figure 10. 9–12 AMAO 1 profile for proficiency scores.

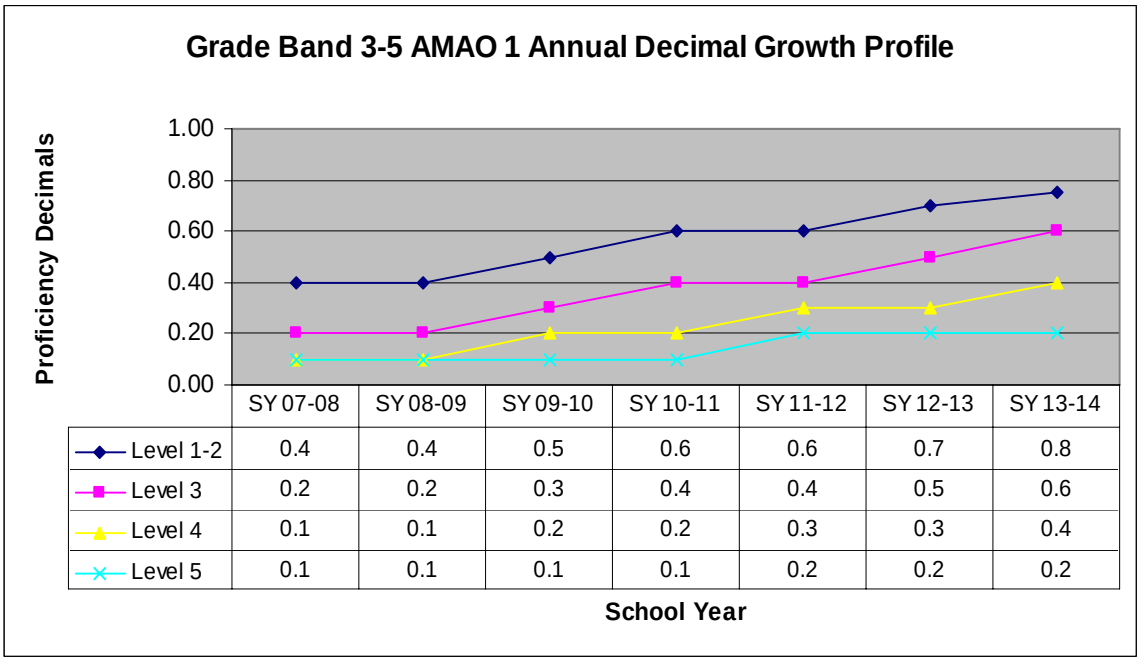
Proficiency Level Decimal AMAO 1 Profiles

Figures 11–14 show AMAO decimal growth profiles.



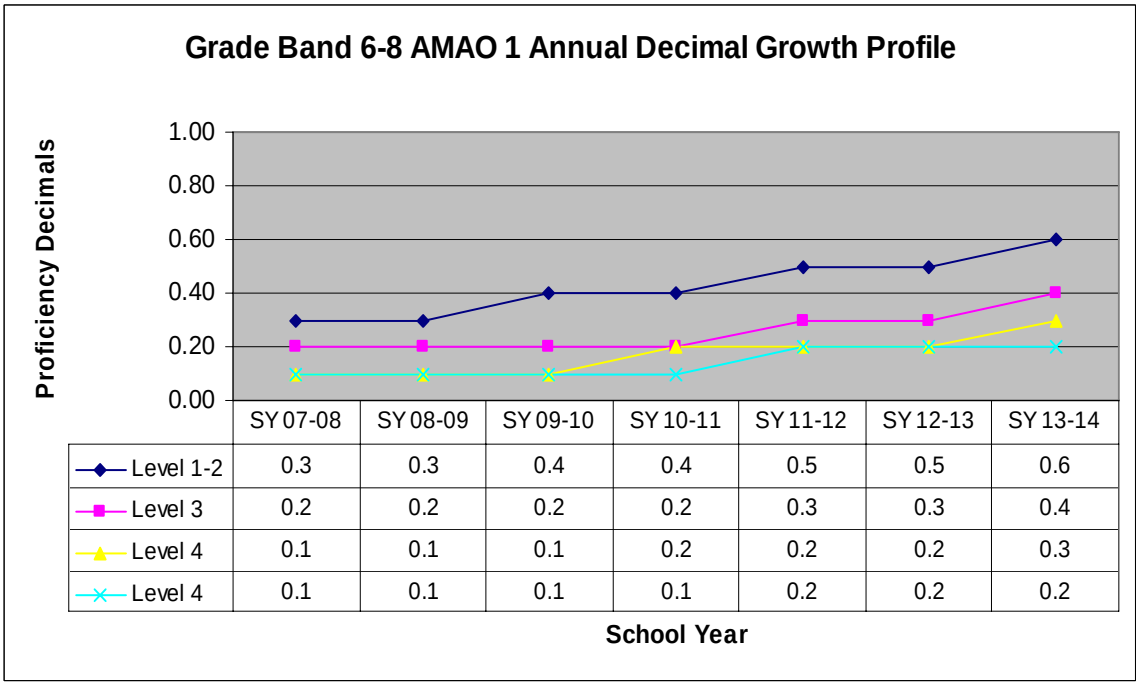
	SY 07-08	SY 08-09	SY 09-10	SY 10-11	SY 11-12	SY 12-13	SY 13-14
Level 1-2	0.6	0.6	0.7	0.8	0.8	0.9	1.0
Level 3	0.3	0.3	0.4	0.5	0.6	0.7	0.8
Level 4	0.2	0.2	0.3	0.4	0.4	0.5	0.6
Level 5	0.1	0.1	0.2	0.2	0.3	0.3	0.4

Figure 11. K–2 AMAO 1 profile for proficiency level decimal scores.



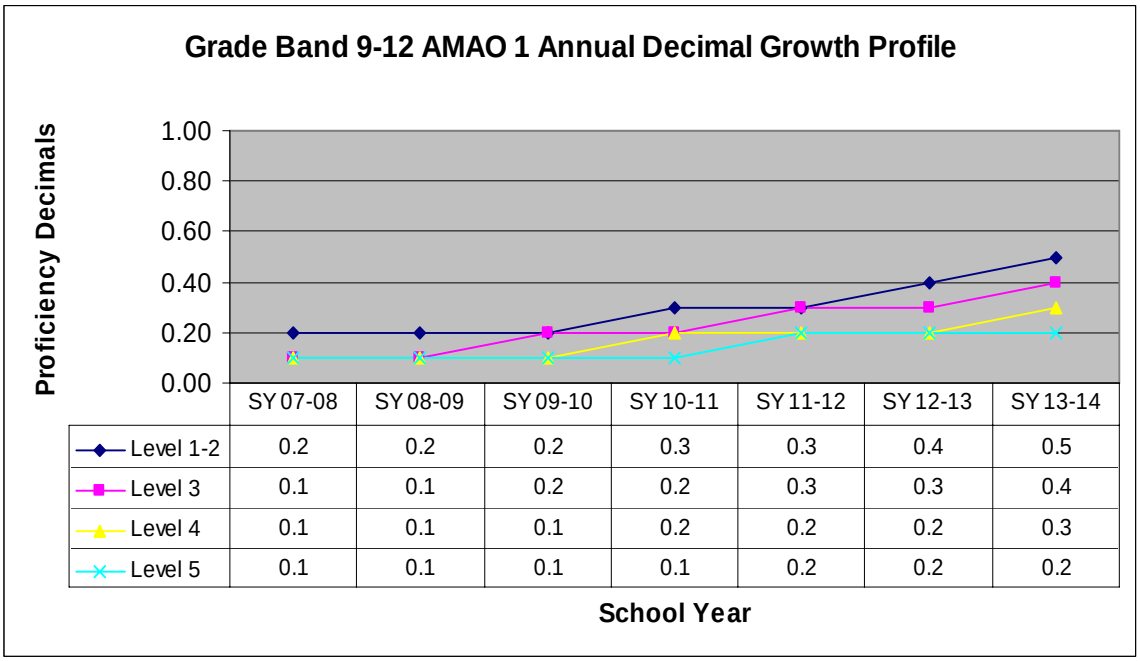
	SY 07-08	SY 08-09	SY 09-10	SY 10-11	SY 11-12	SY 12-13	SY 13-14
Level 1-2	0.4	0.4	0.5	0.6	0.6	0.7	0.8
Level 3	0.2	0.2	0.3	0.4	0.4	0.5	0.6
Level 4	0.1	0.1	0.2	0.2	0.3	0.3	0.4
Level 5	0.1	0.1	0.1	0.1	0.2	0.2	0.2

Figure 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	0.3	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 13. 6–8 AMAO 1 profile for proficiency level decimal scores.

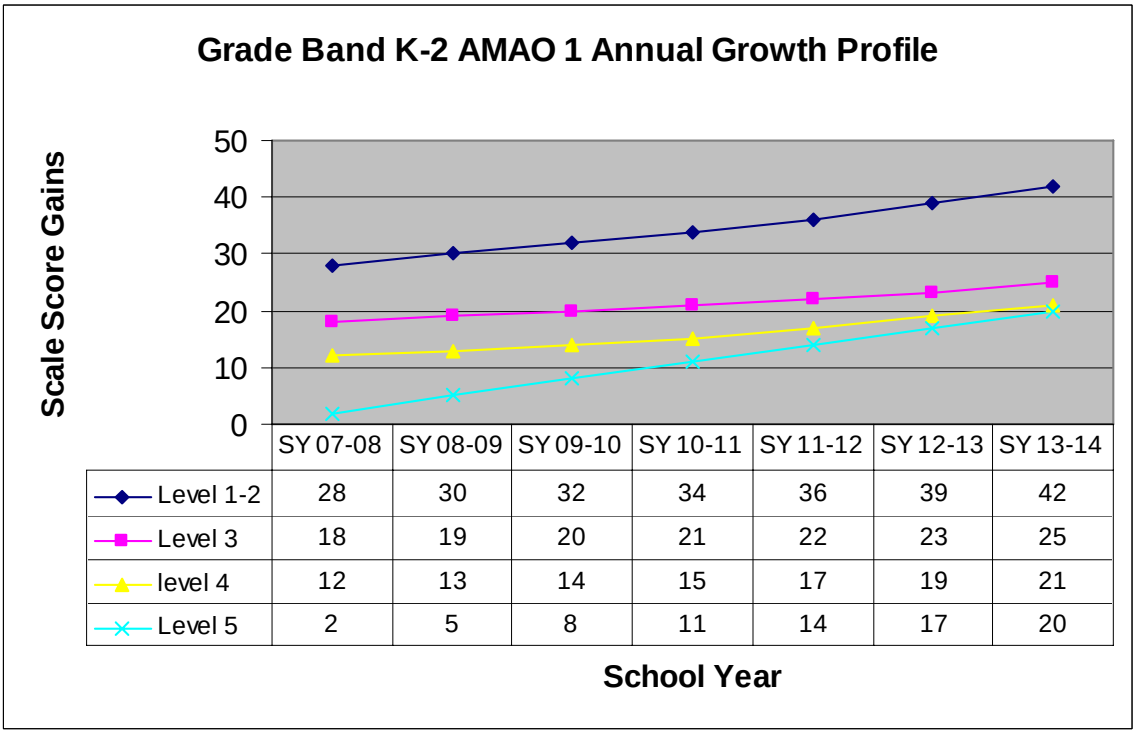


	SY 07-08	SY 08-09	SY 09-10	SY 10-11	SY 11-12	SY 12-13	SY 13-14
Level 1-2	0.2	0.2	0.2	0.3	0.3	0.4	0.5
Level 3	0.1	0.1	0.2	0.2	0.3	0.3	0.4
Level 4	0.1	0.1	0.1	0.2	0.2	0.2	0.3
Level 5	0.1	0.1	0.1	0.1	0.2	0.2	0.2

Figure 14. 9–12 AMAO 1 profile for proficiency level decimal scores.

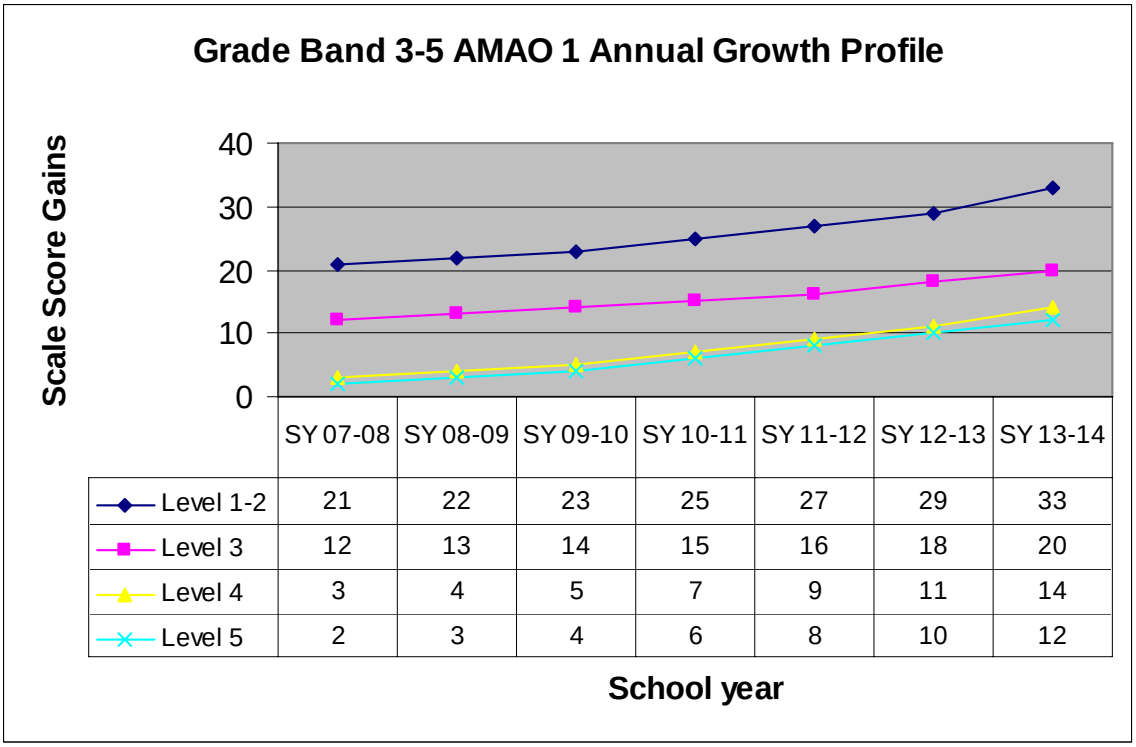
Scale Score AMAO 1 Profiles

Figures 15–18 show AMAO scale score growth profiles.



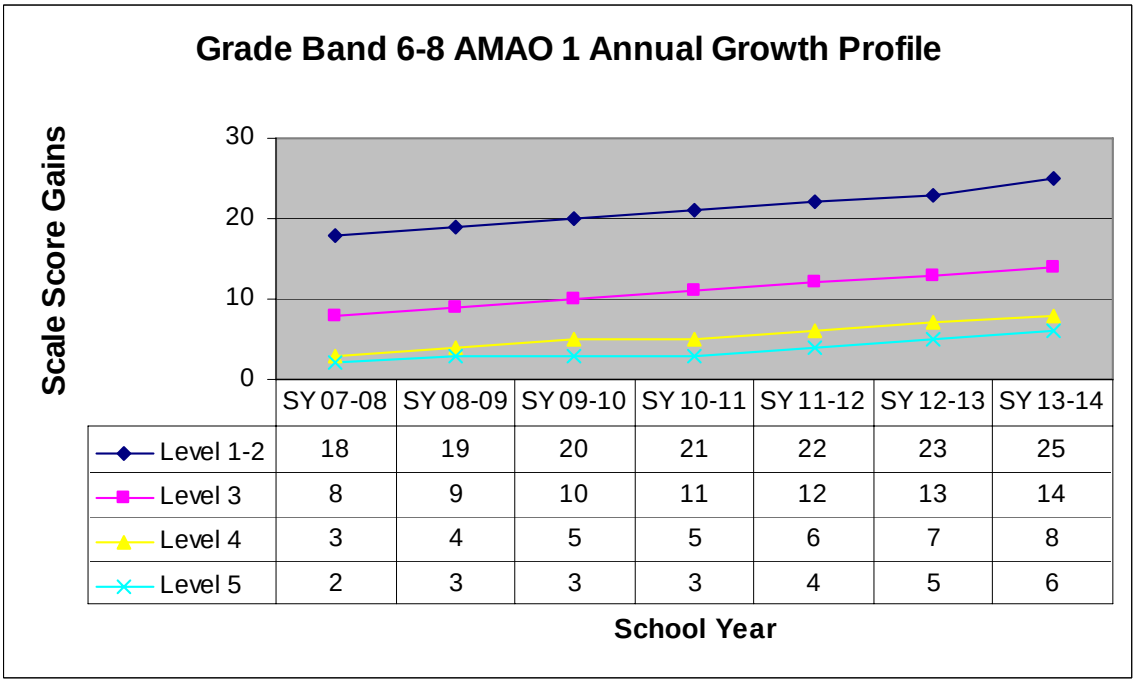
	SY 07-08	SY 08-09	SY 09-10	SY 10-11	SY 11-12	SY 12-13	SY 13-14
Level 1-2	28	30	32	34	36	39	42
Level 3	18	19	20	21	22	23	25
level 4	12	13	14	15	17	19	21
Level 5	2	5	8	11	14	17	20

Figure 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	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 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	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 17. 6–8 AMAO 1 profile for scale scores.

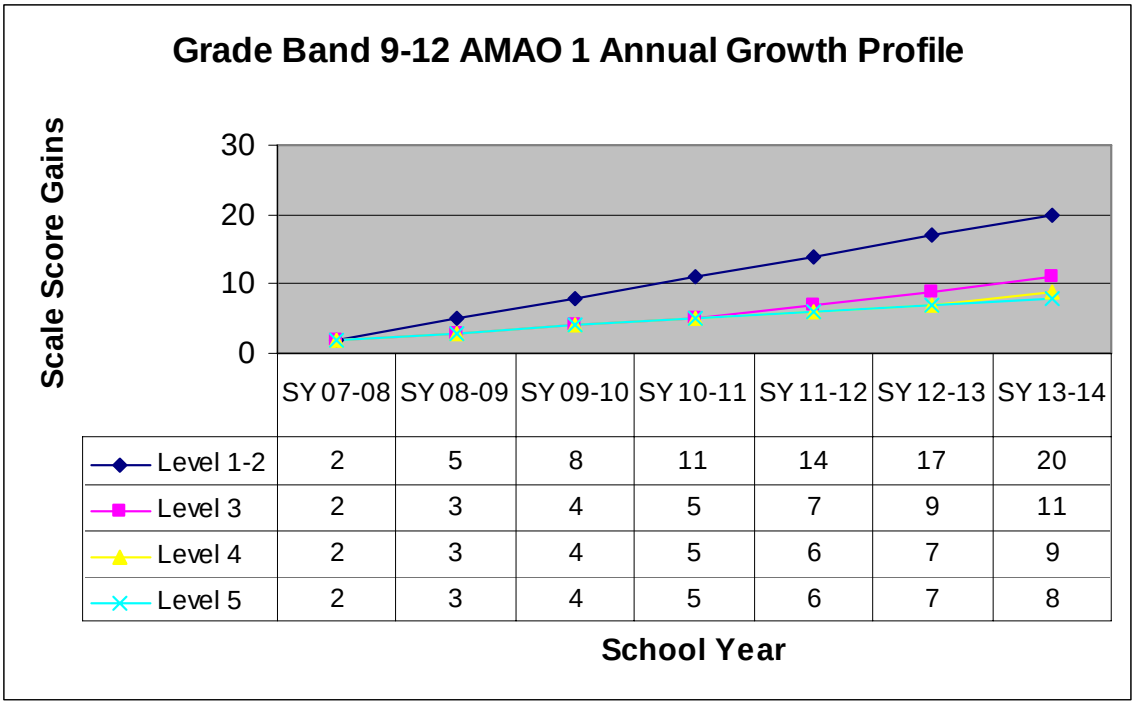


Figure 18. 9–12 AMAO 1 profile for scale scores.

Summary

There are two final points to consider when setting AMAO 1 criteria. First, as previously noted, 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. We strongly suggest examining available state data before adopting grade band and/or proficiency level distinctions. 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, student rankings and distributions might be used to set starting and ending points. But this strategy is problematic, since the aggregate variance in growth at the state level is typically far greater than that at the district level. Basing growth expectations on students is likely to result in expectations that are set too high. 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 AMAO 2: Attainment of English Language Proficiency

Linquanti and George (2007) identified five decisions that need to be made to establish AMAO 2 targets. They are:

1. Define the English proficiency 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, the decision points for AMAO 2 are very similar to those for AMAO 1. The critical differences lie in the first two points, and that is where we turn next.

Define English Language Proficiency

What is English language proficiency in the K–12 school context? This question has been hotly debated for years in the fields of bilingual education, teaching English to speakers of other languages, and applied linguistics (Cummins, 1983; Collier, 1995; Hakuta, Butler, & Witt, 2000). Some definitions focus on students' ability to manage academic English contexts (e.g., Collier, 1995). Others take a more pragmatic view, looking at how students perform or are predicted to perform on standardized state assessments (e.g., Hakuta et al., 2000). Under federal law, a *limited English proficient* (LEP) student is an enrolled, school-aged child:

- Whose native language is not English, and

- Whose difficulties in English deny him or her (a) the ability to perform proficiently on the state's achievement test, (b) the ability to successfully participate in classes in which only English is spoken, and (c) the opportunity to participate fully in society (20 U.S.C. § 7801(25)).

Following from this definition, an *English proficient* student could be said to be one who has the ability to perform proficiently on state tests, successfully participate in classes in which only English is spoken, and fully participate in English-speaking 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 have empanelled state (sometimes national) stakeholder experts to develop standards and English language proficiency expectations (see Cizek, 2001; Hambleton & Pitoniak, 2006). We suggest following a similar strategy in establishing English proficiency as it relates to AMAO 2—specifically:

- 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 with ACCESS scores and determine what level of performance on ACCESS predicts success on the reading and math assessments.

- Identify ELL students who participate successfully in class without substantial English language support and identify which ACCESS proficiency level(s) they represent.

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 help establish meaningful, attainable, and challenging expectations.

Empanel relevant stakeholders. A careful reading of Linquanti and George (2007) reveals that California took an approach similar to that advocated above in establishing its AMAO 2 expectations. Relevant stakeholders might include parents, teachers, administrators, and interested parties who are familiar with and have worked with ELL students and the state's English language proficiency standards and expectations. Panelists from stakeholder groups could be provided with information from the second two activities listed above when they set AMAO 2 levels. As with standards-setting activities, this activity could be conducted in several rounds until agreed-upon levels are established.

Correlate the state's test data. Some sort of correlation or regression analysis should be considered when examining the relationship between a state's content assessments and ACCESS. For example, one could conduct biserial correlations between the state's proficiency levels in reading and mathematics and ACCESS scale scores at each grade band. One might also consider regressing ACCESS proficiency scores to the state's reading and mathematics scores. In both cases, one would be looking for the point at which ACCESS scores cease being predictive or begin to have lower correlations—that is, the point at which language limitations no longer predict academic performance, including performance on the state's achievement tests. English proficiency could then be said to represent the point at which the distribution of ELL students on the state's achievement tests closely matches that of first language English speakers.

Identify ELL students who participate successfully without substantial support. 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.

Determine Cohort of ELLs for Analysis

There are two potential methods for determining which students to include in the AMAO 2 cohort: (a) estimate the time a student at a particular grade band and proficiency level would take to become English proficient or (b) identify students at different proficiency levels who would be included in a cohort from which English proficient percentages would be calculated. We discuss each of these methods below.

Method 1

If states determine cohorts based on time, they must first statistically model the time required to become English proficient based on a student's 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 students at lower proficiency levels grew faster than older students and students at higher proficiency levels, and the rate of growth declined as students moved up proficiency levels. Thus, nonlinear models would need to be fitted.

Next, states would need to create a matrix of expected times based on students' grade bands and proficiency levels. States would then track students over time, keeping in mind where they started and where they currently are, relative to their grade and proficiency level, and determining the percentage of students who are on track.

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, the nonlinear models needed to meaningfully predict student growth require sophisticated statistical analyses that state departments of education, especially Title III directors and their staff, may not have the capacity to perform or explain to districts, schools, and teachers. Second, using this method requires a significant amount of record keeping that may prove challenging for states. In particular, the approach requires a state longitudinal student identification system that can track the grades and proficiency levels at which students began their English language instructional program, their current standing in the program, and the point at which they exited. While most districts have this capacity, many states may not.

Method 2

An alternative would be identify students at different proficiency levels to include in AMAO 2 cohorts. This is the approach Linquanti and George (2007) took in developing an accountability system for the California State Board of Education, and it is the approach we suggest here as well. To apply this method, one must first assign cohorts of students with two data points included in the proficiency calculation. According to Linquanti and George (2007, p. 8), the California board decided to include in the ELL cohort students who met any of the following criteria (note that 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 the Early Advanced and Advanced level but were not English proficient (based on subskills) in the prior year;
- ELLs who were at the Beginning or Early Intermediate level in the prior year and who had first enrolled in U.S. schools at least 4 years earlier; and
- ELLs who were at the Beginning or Early Intermediate level in the prior year, who had been in U.S. schools *less than* 4 years, and who reached the English proficient level in the current year.

Two important questions were answered in devising these cohort criteria. First, at what level should students be included in the English proficient formula for districts? In California's case, that level was Intermediate and above. 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 and in U.S. schools less than 4 years (normally excluded) but who reached proficiency were also to be included. In determining cohort assignment using ACCESS data and WIDA proficiency levels, we suggest using this same procedure.

In examining student proficiency level distributions in the three-state sample and running a variety of cohort combinations, we found that a WIDA proficiency level of 2.5 seemed to result in percentages similar to those identified by Linquanti and George (2007). Furthermore, over a 3-year period in the three-state sample, less than 3% of students 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 English proficiency (at 5.0) in 3 years. In this case, as with Linquanti and George's observations, 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 prior year;
- ELLs who were at a WIDA proficiency level of less than 2.5 in the prior year and who first enrolled in U.S. schools 4 or more years ago; and
- ELLs who were at a WIDA proficiency level of less than 2.5 in the prior year, who had been in U.S. schools *less than* 4 years, and who reached the English proficient level in the current year.

As stated earlier, each state must address the questions of when students should start being counted and how long lower level students should be excluded from the English proficient calculation. The above cohort criteria should be a good place to begin discussing theses issues. *It is important to note that AMAO cohort definitions are still under discussion by the U.S. Department of Education and that states may be required to include all ELLs in AMAO 2 calculations.*

Determine Starting Point for AMAO 2 Targets

To establish a starting point for AMAO 2 targets, Linquanti and George (2007) applied California's AMAO 2 cohort criteria to districts and rank-ordered districts with 25 or more ELL students based on the percentage of students attaining English proficiency. The AMAO 2 starting point was then set at the percentage of English proficient students in the district that was at the 20^{th} percentile of the state's distribution.

For illustrative purposes, we adopted a WIDA proficiency level of 5.0 as the English proficient point. Like Linquanti and George, we rank-ordered districts, but we set the minimum number of students at 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 20^{th} percentile corresponded to 18% 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 Ending Point for AMAO 2 Targets

Title III requires schools to annually increase the numbers of students attaining English proficiency. Fortunately, the legislation does not require 100% of students to achieve proficiency by the end point, a goal that would be unattainable. Where should the end point be? Three alternatives were provided to California policymakers: the 60^{th}, 75^{th}, and 90^{th} percentiles. California chose the 75^{th} percentile, which corresponded to 46% of students in a district attaining

English proficiency for AMAO 2. Using the WIDA three-state data set, the 75th percentile would correspond to 41% of students in a district attaining English proficiency.

Determine Annual Rate of Growth from 2005 to 2014

Once we obtain starting and ending points, we can calculate an annual growth rate. Assume a WIDA level of 5.0 as English proficient, which would result in a starting point of 18% and an ending point of 41%. Figure 19 plots a 9-year timeline of annual growth, from 2005 to 2014. Again, this figure is for illustrative purposes, as each state may choose different objectives and timelines. In Figure 19, we chose the 2005–2006 school year as the starting point because all students were included in the three-state sample that year. Subsequent years had longitudinal rather than census data and would likely have had some bias. When states establish their starting points, they should begin with the 2006–2007 school year. Conceptually, however, the process of establishing annual growth rates will be the same.

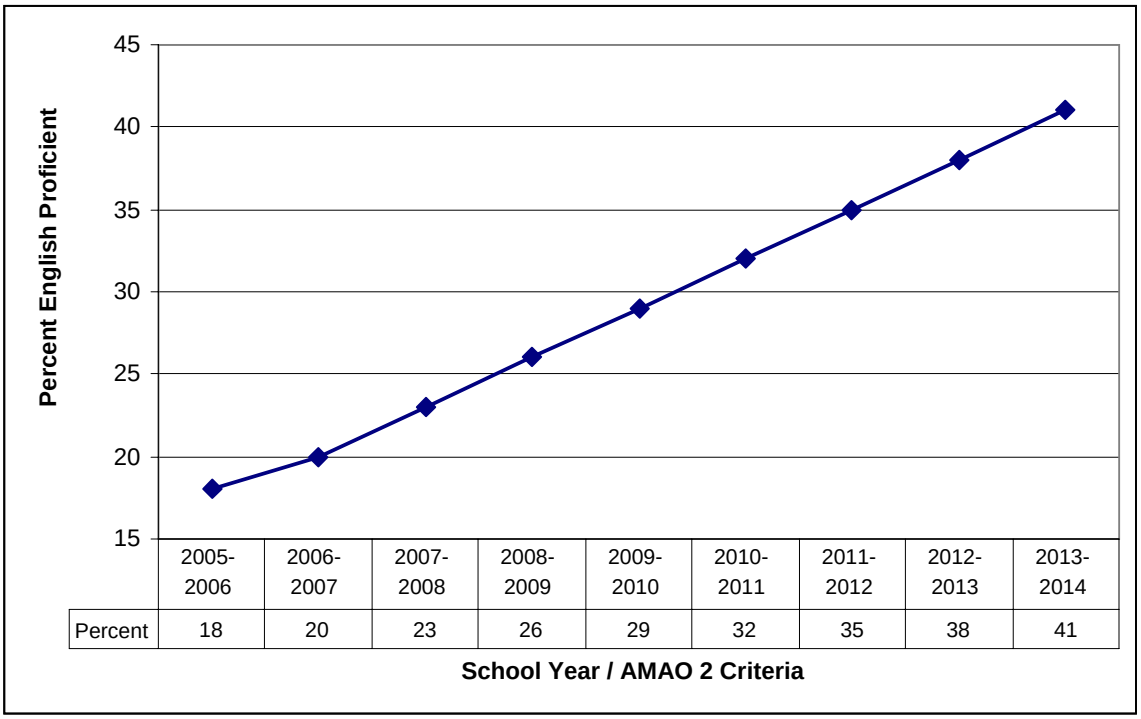


Figure 19. Example of AMAO 2 targets.

As one can see in Figure 19, with 2005–2006 as a starting point, districts would be required to have 23% or more of their students in cohort groups reach English proficiency in the 2007–2008 school year.

One concern with the AMAO 2 targets presented above might be the difference in the percentage of English proficient students at particular grades. If younger students progress faster, districts that service predominantly older students—e.g., the "unified" districts that serve only high school students in many WIDA member states—might be disadvantaged by having one criterion apply across all grade bands. To address this concern, it may be necessary to create grade bands 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. We did this with the current three-state data set and found little difference between the K–5 and 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 AMAOs in English as required under NCLB. We also hope that this paper promotes further dialogue within WIDA states and beyond about the complexities involved in setting challenging yet reasonable growth targets for this diverse group of students we refer to as ELLs.

AMAO 1: Findings and Recommendations

Three possible metrics are available for setting AMAO 1 criteria with the ACCESS assessment: proficiency levels, proficiency level decimals, and scale scores. For all metrics, we found that students made different progress in their language development (based on ACCESS scores) at different grades and proficiency levels, such that students at lower grades/proficiency levels progressed faster than students at higher grades/proficiency levels. This finding suggested the need to tailor growth expectations to students' initial grades and proficiency levels.

Tables 11–13 summarize key AMAO 1 findings based on the three-state WIDA data set.

Table 11
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 12
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 13
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 questions to be answered in establishing AMAO 2 criteria, two are critical: “What is English proficiency?” and “Which cohorts of students should be considered when calculating the percentage of students who are English proficient in districts?”

We suggested that the following activities be considered when setting English proficiency 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 with ACCESS scores and determine what level of performance on ACCESS predicts success on the reading and math assessments.

3. Identify ELL students who participate successfully in class without substantial English language support and identify which ACCESS proficiency level(s) they represent.

We identified 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 prior year;

2. ELLs who were at a WIDA proficiency level of less than 2.5 in the prior year and who first enrolled in U.S. schools 4 or more years ago; and

3. ELLs who were at a WIDA proficiency level of less than 2.5 in the prior year, who had been in U.S. schools *less than* 4 years, and who reached the English proficient level in the current year.

Alternative Approaches to Determining AMAO 1 Expectations

Two alternative approaches to establishing AMAO 1 expectations might also be considered. They differ from the methods used by Linquanti and George (2007), 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 the letter.

Growth Normalization

The guiding principle in this paper is "lower is faster, higher is slower." In line with this principle, we applied different expectations to different grade bands and proficiency levels. Another approach, called *growth normalization*, would create a common metric that "normalizes" or places different growth expectations on 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, i.e., 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.

Thus, g_i is 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 performance. 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 to the three-state sample with promising results. The benefit of this procedure is that it results in one specific growth expectation score that reflects expected gain across grades and proficiency levels.

In setting a starting point, we might choose 1 or 1½ standard deviations below the initial mean, and the ending point might be 1 or 1½ standard deviations above the initial mean. A drawback of this approach is that we are losing information about individual student progress as a result of normalizing growth. We are comparing current student progress to past progress, 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. One could transform g_i 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 alternative approach to identifying AMAO 1 expectations is to fit a mixed linear model to available test data. This approach provides estimates of the differing growth trajectories for different proficiency levels in a grade band. In essence, we are predicting the nature of student growth. A necessary requirement for this model is an interval, vertically scaled scoring metric. This procedure cannot be applied to proficiency levels or proficiency level decimal scores. The following model illustrates how we might estimate growth projections:

$$\text{Level 1}$$

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

$$\text{Level 2}$$

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

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

$$\text{Combined Model}$$

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

where in the combined model, y_{ti} represents student i's predicted ACCESS composite score; β_{00} is the predicted initial composite score; 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. Figure 20 displays SAS code for this type of analysis.

```
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;
```

Figure 20. SAS code for mixed-model analyses.

Once growth is estimated, predicted gains by grade band and proficiency level can be obtained. To set starting and ending points, standard errors associated with growth estimations can be applied. For example, the starting point could be set at 2 standard errors below the predicted growth estimate, and the ending point at 2 standard errors 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 unless it controls for students who exit early. 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)—is necessary to correctly estimate student growth.

We conducted preliminary analyses using a predictive growth model using the three-state data set. Our results suggest that this procedure has promise; however, more research is necessary to justify its use. Value-added modeling techniques could also be applied using this procedure. For more detail about value-added models, see Wainer (2004).

References

Birdsong, D. (Ed.). (1999). *Second language acquisition and the critical period hypothesis*. Mahwah, NJ: Lawrence Erlbaum.

Cizek, G. J. (Ed.). (2001). *Setting performance standards: Concepts, methods, and perspectives*. Mahwah, NJ: Lawrence Erlbaum.

Collier, V. (1987–1988). The effect of age on acquisition of a second language for school. *New Focus, 2*. Retrieved February 14, 2008, from http://www.ncela.gwu.edu/pubs/classics/focus/02aage.htm

Collier, V. (1995). Acquiring a second language for school. *Directions in Language & Education, 1*(4). Retrieved February 14, 2008, from http://www.ncela.gwu.edu/pubs/directions/04.htm

Cummins, J. (1983). Language proficiency and academic achievement. In J. Oller (Ed.), *Issues in language testing research* (pp. 108–130). Rowley, MA: Newbury House.

Doughty, C. J., & Long, M. H. (2003). The scope of inquiry and goals of SLA. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 3–15). Malden, MA: Blackwell.

Hambleton, R. K., & Pitoniak, M. J. (2006). Setting performance standards. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 433–470). Westport, CT: Praeger.

Hakuta, K., Butler, Y. G., & Witt, D. (2000). *How long does it take English learners to attain proficiency?* (Policy Report 2000-1). Santa Barbara: University of California Linguistic Minority Research Institute. Retrieved February 14, 2008, from http://www.lmri.ucsb.edu/publications/00_hakuta.pdf

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica, 47*(1), 153–161.

Hyltenstam, K., & Abrahamsson, N. (2003). Maturational constraints in SLA. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 539–588). Malden, MA: Blackwell.

Krashen, S. (1987). *Principles and practice in second language acquisition.* New York: Prentice-Hall International.

Linquanti, R., & George, C. (2007). 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 nation: Current status and future practice* (pp. 105–118). Davis: University of California, Davis, School of Education. Retrieved February 14, 2008, from http://education.ucdavis.edu/research/ELP_Report.pdf

Long, M. H. (2003). Stabilization and fossilization in interlanguage. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 487–535). Malden, MA: Blackwell.

McLaughlin, B. (1984). *Second-language acquisition in childhood: Vol. 2. School-age children* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

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.

No Child Left Behind Act of 2001, Pub. L. 107–110, 115 Stat. 1425 (2002).

Scarcella, R. (2003). *Academic English: A conceptual framework* (Technical Report 2003-1). Santa Barbara: University of California Linguistic Minority Research Institute. Retrieved February 15, 2008, 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.

Wainer, H. (Ed.). (2004). Value-added assessment special issue [Special issue]. *Journal of Educational and Behavioral Statistics, 29*(1).

WIDA Consortium. (2007). *ACCESS for ELLs® interpretive guide for score report.* Madison, WI: Author. Retrieved February 12, 2008, from http://www.wida.us/assessment/ACCESS%20Interpretive%20Guide07.pdf

Appendix A
Sample of WIDA State AMAO Policies

State	Grades	AMAO 1 definition	Title III met if…	AMAO 2 definition	Title III met if…
Alabama	All	Move to next cohort annually (I=1.0-2.0, II=2.1-2.8, III=2.9-3.5, IV=3.6-4.1, V=4.2-4.7)	Now: 60% 2014: 80% growth target (+ 2.5%/year)	Level 4.8	100% of students attain English proficiency in 5 years or less
Illinois	All	.5 level increase in 1 of the 4 domains	85% of students (if N>30)	Cutoff score identified by ISBE, currently Level 4.0	10% of students meet this score
Maine	K-8	90% of LEPs advance 1 level from Levels 1-3 90% of LEPs advance within at least Levels 4-5	See left	Level 6 in each domain	Not specified
	9-12	80% of LEPs advance 1 level from Levels 1-3 80% of LEPs advance within at least Levels 4-5			
New Jersey	1-2 3-5 6-8 9-10 11-12	+19 points on ACCESS for ELLs (scale points) +14 points +12 points +7 points +4 points	2003/2004: 50% of students meet expectations, increases by 5% yearly to 100% in 2014	Level 3 (K), Level 4.5 (1-12) "if multiple criteria support the decision"	2003/2004: 50% of students meet expectations, increases by 5% yearly to 100% in 2014
Rhode Island	All	.5 district-wide average growth score	See left	Level 4.5	40% of students in LEP 3 years must be proficient
Vermont		Previous year's composite English language proficiency level	50% (tentative)	Level 5	If meets function of the percent of students who would be proficient if all growth targets for making progress are met
		<2 / 2-3.5 / >3.5			
	K	1 / 1 / N/A			
	1-2	0.8 / 0.8 / 0.5			
	3-5	0.8 / 0.5 / 0.3			
	6-8	0.5 / 0.3 / 0.2			
	9-12	0.5 / 0.2 / 0.2			
Wisconsin	All	90% of Level 1 students progress by 1 level 90% of Level 2-4 students progress by .5 level	See left	90% of Level 5 students become fully proficient	See left