WIDA™
CONSORTIUM

World-Class Instructional Design and Assessment
Wisconsin Center for Education Research
University of Wisconsin-Madison
1025 West Johnson Street, MD #23
Madison, WI 53706

608.263.4216 • 866.276.7735 • www.wida.us

TO: MR. RICHARD L. SMITH, LEP PARTNERSHIP

FROM: THE WIDA CONSORTIUM

SUBJECT: COMMENTS ON FEDERAL REGISTER "NOTICE OF PROPOSED INTERPRETATION" REGARDING AMAOS—RELEASED MAY 2, 2008

DATE: MAY 30, 2008

CC: KATHRYN DOHERTY, WIDA SEAS IN 19 MEMBER STATES

The Office of English Language Acquisition (OELA), U.S. Department of Education released a Federal Register[1] with the purpose of providing a notice of interpretation on issues related to establishing and implementing annual measurable achievement objectives (AMAOs) on May 2, 2008 (hereafter *Notice*). This *Notice* was released with the intent of collecting comments regarding these interpretations.

H. Gary Cook, Embedded Researcher for WIDA from the Valued Added Research Center (VARC), is the primary author of this response which has been shared with the WIDA Consortium membership. Tim Boals, WIDA Executive Director, is the secondary author. We do not affirm that all our member states agree with all comments that follow, but believe that collectively the concerns within this memo do represent concerns expressed by our state partners and certainly the WIDA Consortium leadership, who have invested considerable time and energy looking at longitudinal data on English language learner progress, in an effort to more fully appreciate the complexity of these issues. The WIDA Consortium is also encouraging individual states within our consortium to draft responses as we believe the issues covered in the interpretation have far reaching implications.

As an embedded research scientist for WIDA, H. Gary Cook has been exploring areas of research that the WIDA Board deems important. One of his first assignments was to provide guidance on establishing AMAOs using WIDA's ACCESS for ELLs® assessment data. That research was conducted and a report was published.[2] He has consulted with several states, interacted with OELA and several nationally recognized researchers on AMAOs. Tim Boals, Executive Director of the WIDA Consortium based at the Wisconsin Center for Education Research, regularly meets and discusses AMAO related issues and broader ELL policy concerns with WIDA SEAs. His background is in curriculum/language development and educational policy for English language learners.

WIDA is concerned about several of the interpretations outlined in this notice and feel obligated to share these concerns with you. What follows are comments based on our research, discussions at WIDA and with individual SEAs in 19 states on each interpretation for which we have a concern.

---

[1] Title III of the Elementary and Secondary Education Act of 1965 (ESEA), as Amended by the No Child Left Behind Act of 2001 (NLCB), Notice of Proposed Interpretation, Fed. Reg. 24266-24272 (May 2, 2008) (to be codified at 34 CFR. pt.79).

[2] See www.wcer.wisc.edu/publications/workingPapers/Working_Paper_No_2008_02.pdf

WIDA EMBEDDED RESEARCHER

TO: TIM BOALS, PH.D.

FROM: H. GARY COOK, PH.D

SUBJECT: COMMENTS ON FEDERAL REGISTER "NOTICE OF PROPOSED INTERPRETATION" REGARDING AMAOS—RELEASED MAY 2, 2008

DATE: 5/23/2008

CC: M. ELIZABETH CRANLEY, PH.D.

As you know, the Office of English Language Acquisition (OELA), U.S. Department of Education released a Federal Register[1] with the purpose of providing a notice of interpretation on issues related to establishing and implementing annual measurable achievement objectives (AMAOs) on May 2, 2008 (hereafter *Notice*). This *Notice* was released with the intent of collecting comments regarding these interpretations. The deadline for submitting comments is June 2, 2008. I intend on submitting comments to OELA prior to June 2nd but thought it might be helpful to share my concerns with your first, thinking that other states might have similar issues. Please feel free to share this memo with others.

My role as a research scientist for WIDA is to explore areas of research that the WIDA Board deems important. One of my first assignments was to provide guidance on establishing AMAOs using WIDA's ACCESS for ELLs® assessment data. That research was conducted and a report was published.[2] In various capacities, I have been working on issues associated with AMAOs for the better part of a year now. I have consulted with several states, interacted with OELA and several nationally recognized researchers on AMAOs, so I feel that I have some insight regarding the process of their establishment. I am concerned about several of the interpretations outlined in this notice and feel obligated to share my concerns with you. What follows are my comments on each interpretation for which I have a concern. It would be most helpful to have a copy of this Notice with you as your read my comments; thus, I've attached a copy of the Notice in the email message with this memo.

1. ANNUAL ELP ASSESSMENTS OF LEP STUDENTS

***Interpretation**: The Secretary proposes to interpret section 3113(b)(3)(D) to require that all LEP students be assessed annually with an assessment or assessments that measure each and every one of the language domains of speaking, listening, reading, and writing.*

[1] Title III of the Elementary and Secondary Education Act of 1965 (ESEA), as Amended by the No Child Left Behind Act of 2001 (NLCB), Notice of Proposed Interpretation, Fed. Reg. 24266-24272 (May 2, 2008) (to be codified at 34 CFR. pt.79).

[2] See www.wcer.wisc.edu/publications/workingPapers/Working_Paper_No_2008_02.pdf

I have not particular concern with this interpretation. It eliminates the provision of "banking" domain test scores. That is, if students receive a proficient score on a domain test they would no longer be required to take that test again. This provision is removed. I concur with this interpretation. It could be that students who receive a "proficient" score on a domain test (say listening) may indeed be proficient in reading. But, it could also be plausible that they are not "proficient" but scored high because they guessed rightly or were just involved in a discussion of a topic in class that happened to be on the English language proficiency exam. Standard error of measure for students' scores occurs in both directions, i.e., students score higher than their true score or score lower.

2. USE OF ANNUAL ELP ASSESSMENT SCORES FOR AMAOS 1 AND 2

Interpretation for AMAO 1*: With regard to AMAO 1, the Secretary proposes to interpret Title III to allow States to base their student performance expectations and accountability (i.e., AMAO 1) targets for progress on assessment results derived from either (1) separate student performance levels or scores in each of the language domains or (2) a single composite score or performance level derived by combining performance scores across domains, so long as such a composite score can be demonstrated to be a valid and effective measure of a student's progress in each of the English language proficiency domains. The Secretary also proposes to interpret Title III to allow States to determine their AMAO 1 targets based on progress in one or more of the language domains, rather than requiring student progress separately in each and every one of the language domains, so long as the targets provide for meaningful progress toward attaining English language proficiency.*

My chief concern here is the provision to establish AMAO 1 criteria based on one or two domain scores. In the development of standards associated with academic English language proficiency, states and consortia have inexorably connected domains to assess their particular operationalized form of academic language proficiency (ALP). The notion of ALP espoused by many (if not most) states involves academic discourse, i.e., the discourse of English language arts, mathematics, science, and social studies, which includes productive (speaking & writing) and receptive (listening & reading) skills. Isolating one domain limits the generalization of how students grow in the discourse of the classroom. I would strongly urge states to establish AMAO 1 criteria using all domain scores, be they weighted or unweighted. Meaningful progress, in my view, involves all the domains of discourse not just one or two components.

Interpretation for AMAO 2*: With regard to AMAO 2, attaining English language proficiency, the Secretary proposes to interpret Title III to allow States to base their student performance expectations and accountability targets for attainment on assessment results derived from either (1) separate student performance levels or scores in each of the language domains or (2) a single composite score or performance level derived by combining performance scores across domains, provided that such a composite score can be demonstrated to be a valid and effective measure of a student's proficiency in each of the English language proficiency domains. In setting student performance expectations and accountability targets for attaining proficiency in English (AMAO 2), it is the Secretary's proposed interpretation of Title III that a LEP student must score proficient or above in each and every language domain required under Title III in order to be considered to have ''attained proficiency'' on a State's ELP assessment. If a State's ELP assessment generates a composite score, the State would have to demonstrate that an overall proficient ELP score represents proficiency in all domains for students served by Title III.*

In principal, I have no argument with this interpretation. There are several issues to consider, however. First, if a state chooses to establish AMAO 2 criteria based on separate domain scores, the probability of identifying a district as not meeting this criterion increases—merely based on the number of decisions that must be made (4 instead of one). Second, a negative consequence of this interpretation might be that states chose to lower the expected AMAO 2 criterion to minimize the number of identified districts, regardless of whether they use four domain scores or one composite score. Third, essentially this interpretation imposes a conjunctive decision making model on states' establishment of AMAO 2. By conjunctive, I mean that all domain scores must be proficient before a student is considered proficient for AMAO 2 purposes. An alternative view would be a compensatory model. A compensatory model would have one composite score as the criterion and domain scores would fluctuate as long as the overall composite was above the criterion. This clearly is not allowed based on this interpretation. How would a model that sets a composite proficiency level score at 5.0 (using WIDA scores), lets say, and no domain scores below 4.5 be interpreted? That is, students would have to have all domain scores at or over 4.5 and a total composite score of 5.0 or greater. With this model, there are compensatory features, but there is a floor below which students cannot go to be considered proficient. In my view, this WOULD BE an acceptable model under this interpretation, and in fact, based on OELA's interpretation one that I would recommend.

3. STUDENTS INCLUDED IN TITLE III ACCOUNTABILITY

Interpretation*: The Secretary proposes to interpret Title III to require that all LEP students served by programs under Title III be included in all AMAO targets, calculations, and determinations. In addition, the Secretary proposes to interpret Title III, consistent with Title I, as requiring all LEP students attending a public school within a State or subgrantee's jurisdiction—not only those LEP students served by Title III programs—to be included in targets, calculations, and determinations for purposes of determining whether a State or Title II subgrantee meets AMAO 3.*

I have major concerns about this interpretation. I will limit my comments here to AMAO 2 specifically. By definition, an English language learner (federally-limited English proficient-LEP) 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)).

By definition, AMAO 2 is about students who have gained English language proficiency. Researchers have found that the amount of time required to reach English proficiency is anywhere between 3 and 7 years.[3] How shall we treat students who have just arrived in the U.S.

[3] 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 May 16, 2008, from http://www.lmri.ucsb.edu/publications/00_hakuta.pdf

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

and are at the lowest proficiency levels? Given available research, students at the lowest proficiency levels i.e., in the early stage of learning English, are not expected to obtain English proficiency in just one year—or two years for that matter. Yet this interpretation requires all students be included in calculation of AMAO 2, even those who are not expected to obtain proficiency in the next year. Does that make sense? In my view, it does not. Certainly, by allowing exclusions, there is the potential for "gaming the system." This is not a desirable outcome for it potentially sustains bad programs. But requiring ALL students to be included in AMAO 2 may potentially misidentify good programs.

A majority of states' AMAO 2 policies are based on the percentage of students gaining proficiency from last year to this. With a premise that AMAO 2 is based on percentages gaining proficiency from last year, let us look at an example. If District A has 100 students, 10 of whom were new last year and borderline proficient and District B has 100 students, 10 of whom were new last year and at the lowest proficiency level, how would AMAO 2 calculations be affected? Let us further imagine that 20 of the remaining "non-new" students in District A became proficient and 30 from District B became proficient this year. Thus in both District A and District B, 30% gained proficiency this year. Based on the "all-in" interpretation discussed here, both districts would be judged equally. But are they? Could it be that District B actually has better programs, but the influx of new low level students (not expected to be proficient in one year) suppressed this? Not accounting for what we know about child language acquisition when establishing the AMAO 2 criterion may misidentify districts. This is not a desirable outcome.

There is a potential compromise. I suggest that students at the lowest proficiency levels (in WIDA states that might be levels 1.0 to 2.5) have their participation in AMAO 2 calculations be weighted or indexed. If low level students obtain proficiency, their status will be fully counted. This way, students who are unlikely to gain proficiency in one year have limited effects on AMAO 2 calculations, but are still all included in AMAO 2 calculations. Let us also say that low level students have a limited timeframe (say 2 years) to be awarded weighted or indexed scores. This would preclude programs from accepting students staying at lower proficiency levels.

Including all students in AMAO 2 calculations without some form of indexing or weighting of the lowest level students, WILL misidentify. I urge states to consider my alternative and enjoin OELA to accept this approach as viable.

4. EXCLUSION OF LEP STUDENTS ''WITHOUT TWO DATA POINTS'' FROM AMAO 1

Interpretation*: The Secretary proposes to interpret the requirement in section 3122(a)(3)(A)(i) of the ESEA to include all LEP students served by Title III in measurements of student progress in English (AMAO 1). This would mean that all such students would have to be*

---

the nation: Current status and future practice (pp. 105–118). Davis: University of California, Davis, School of Education. Retrieved May 16, 2008, from http://education.ucdavis.edu/research/ELP_Report.pdf

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 May 16, 2008] from http://www.wcer.wisc.edu/publications/workingPapers/papers.php

included regardless of whether they have participated in at least two consecutive and consistent annual administrations of an ELP assessment required under section 3113 of the ESEA. Under this proposed interpretation, all LEP students served by programs under Title III would have to be included in AMAO 1 determinations. If a State does not have the requisite two years of data for some LEP students served by Title III in the State, the State would be permitted to propose to the Department an alternative method of calculating AMAO 1. The Department would require that the alternative method for measuring progress under AMAO 1 be based on research on how LEP children acquire proficiency in English and include reliable measures of growth in English language proficiency. Under this proposed interpretation, the Secretary also would allow States to include criteria—in addition to progress on an annual ELP assessment—to be factored into progress determinations for AMAO 1, even for students who have participated in two consecutive administrations of the required annual ELP assessments.

I have great concerns about this interpretation. It is clear from the *Explanation* section of this section that the Department recognizes you cannot calculate gain with only one score. In the WebEx session held on May 7th, 2008, the Department mentioned that 40% of ELLs were not included in AMAO 1 calculations in the past year—based on their reports. Given the transitory nature of ELLs, either by moving to another LEA or exiting programs due to proficiency, it is unclear if what the aforementioned 40% entails. Let us assume that this number is indicative of states' AMAO 1 calculations. Having only 60% of students included in an accountability system is very limiting, especially given the unknown nature of the missing 40%. What then is the remedy? I see four remedies, of which two are offered by *Notice*:

- Use the SEA's or LEA's placement tests as the first measure to calculate growth (given within the first 30 days of a students entry to an LEA)
- Use of other measures of students' progress (e.g., formative assessments, expert judgment)
- Use of the annual ELP assessment
- Fixing students' gain scores for those who do not have two data points.

Regardless of the remedy, each must be "based on research on how LEP children acquire proficiency in English and include reliable measures of growth in English language proficiency (p.24269)." Given that the research on LEP students' language acquisition on English language proficiency assessments is somewhat sparse, this will be a challenge. I will discuss each of the above four points and outline my concerns for each.

The first remedy (offered by the *Notice*) is to use the SEA or LEA mandated placement assessment. This seems to be reasonable until one examines the types of assessments used and their designed purposes. Not all states have placements assessment designed in concert with or aligned to their annual ELP assessment. Thus an alignment between the placement test and the annual ELP assessment would be necessary. Placement assessments that are in place (e.g., W-APT for WIDA states) are screening tools and measure a gross concept of English proficiency—not the fine grained focus of the annual assessments. They provide a basic idea of students' proficiency levels to support program placement. Any misplacement can be dealt with in the classroom and with educators and parents. These placement measures have not been designed to broadly cover all ELP standards and are typically 1/3 to 1/2 the size of the annual ELP

assessment. It is almost certain that the validity evidence necessary to use these measures as metrics for growth with the annual ELP assessment is not available, if it exists at all. The *Standards for Educational and Psychological Testing*[4] outlines the required evidence to assure quality assessments. Many of the listed standards apply to this placement tool, but three are of particular interest here.

Standard 1.4: If a test is used in a way that has not been validated, it is incumbent on the user to justify the new use, collecting new evidence if necessary.

Standard 13.17: When change or gain scores are used, such scores should be defined and their technical properties should be reported.

Standard 15.3: When change or gain scores are used, the definition of such scores should be made explicit, and their technical qualities should be reported.

It is clear from these three standards that substantial validation work is necessary to use scores mentioned as the first remedy. Complying with these standards would place a substantial financial burden on states and would take at least 3 years to complete. Content alignment is a critical feature of validity and certainly these placement instruments would need to align appropriately to ELP standards. This would almost certainly mean longer assessments and more burden on states to fund.

The second remedy is to use other measures of student progress, which the *Notice* terms as "additional criteria" or "additional relevant language acquisition data." These additional criteria, which we interpret to be judgmental, formative-like or benchmark-like metrics, would need to comply with the all above mentioned standards and in our view, two additional standards:

Standard 1.7: When a validation rests in part on the opinion or decisions of expert judges, observers or raters, procedures for selecting such experts and for eliciting judgments or ratings should be fully described. The qualifications, and experience, of the judges should be presented. The description of procedures should include any training and instructions provided, should indicated whether participants reached their decisions independently, and should report the level of agreement reached. If participants interacted with one another or exchanged information, the procedures through which they may have influenced one another should be set forth.

Standard 1.17: If test scores are used in conjunction with other quantifiable variables to predict some outcome or criterion, regression (or equivalent) analyses should be include those additional relevant variables along with the test scores.

For some states, systems that assure the validation of more subjective measures are in place. In many, these systems are not; hence, an increase burden would be realized.

The third remedy is to administer the state's annual ELP assessments twice for incoming students. Most ELP assessments have validation system to assure quality. The gain obtained with this remedy would be consistent with AMAO 1 procedures. There would certainly need to be research examining how students gain on the state's ELP assessment at differing time points

[4] American Educational Research Association, American Psychological Association, and the National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing.* Washington, D.C.: American Educational Research Association.

and gain scores would need to be pro-rated accordingly. But this is doable. However, this remedy would cost more for states because new students would be double tested. Also, many states have a limited number of available forms. Either forms would have to be re-used or more forms created.

The forth remedy could potentially take two forms. First, states could fix growth rates based on expected trajectories of students at specific proficiency levels. For example, we have evidence that students at lower proficiency levels grow faster than students at higher levels (see Cook, et al 2008 footnoted earlier). Also students at lower grades grow faster than students at higher grades. Based on this, states could establish "fixed" growth rates based on state averages for students with only one data point. There are problems with this approach. One would be it assumes that students' grow rates are generally stable across programs, grades and levels. This may not be true. The effect of program or district would be masked accordingly. Second, since students are missing one data point, a state could assume that no growth occurred. This is somewhat punitive and less likely to be an adopted strategy. Validating the first approach would take several time points—3 or more. Some states could not apply right away.

Taken together, it seems that the first or third remedies are most plausible. Whatever choice is made (and there are certainly more that I haven't thought of) more resources are needed to comply with this interpretation. Another point worth considering is why there is such a large exclusion rate for AMAO 1. Are there ways to get greater numbers of students into states' models? How and in what fashion are students excluded from AMAO 1? It seems that this question needs to be addressed as well.

Is it necessary to include all students in AMAO 1 calculations? Certainly AYP formulae do not. Title I's AYP calculations presume full academic year (FAY) status of students. Aren't, new students' growth characteristics most appropriately apportioned to previous—not current—educational experiences, i.e., other LEAs or SEAs? Why then include students' growth in an AMAO 1 model in an LEA that has had little influence on students' previous language acquisition?

I suggest amending this interpretation to allow districts to include students for which they have had reasonable influence over. For discussion, let us presume that means students who have been in district for 6 months or more. In this case, AMAO 1 calculations will have some relevance to the current LEA. As is, this interpretation will likely misidentify LEAs based on the quality of the initial measurement and the level and educational experience of incoming students.

5. ATTAINMENT OF ENGLISH LANGUAGE PROFICIENCY AND "EXITING" THE LEP SUBGROUP

I have no comments on this interpretation.

6. USE OF MINIMUM SUBGROUP SIZES IN TITLE III ACCOUNTABILITY

I have no comments on this interpretation.

7. ALL LEP STUDENTS, ADEQUATE YEARLY PROGRESS, AND AMAO 3

I have no comments on this interpretation.

8. AMAOS AND THE USE OF COHORTS

Interpretation: *With this notice of interpretation, the Secretary proposes to interpret Title III to mean that (a) States may, but are not required to, establish 'cohorts'' for AMAO targets, calculations, and determinations; and (b) States may set separate AMAO targets for separate groups or ''cohorts'' of LEP students served by Title III based only on the amount of time (for example, number of years) such students have had access to language instruction educational programs.*

I am encouraged that this interpretation allows for the use of cohorts and that states may set separate targets for cohorts. I will present data from WIDA three states showing why that is a good idea. **But, I steadfastly oppose limiting cohort use to only the amount of time in "language instruction educational program."** I will present WIDA data showing that this is inconsistent with available trends and will very likely misidentify LEAs. I call on OELA to modify this interpretation such that proficiency levels can be used in AMAO 1 calculations. This interpretation could require years in program be used and allow cohorts to be set up using proficiency levels. Requiring only the use of program years WILL set up a system of misidentification relative to AMAO 1 in WIDA states and possibly in states using other assessments as well.

I provide three examples of ways to calculate AMAO 1 using years in program. The data used in this sample come from three WIDA states using ACCESS for ELLs® data. In particular, data are from the 2004-2005 and 2005-2006 school years with a total sample size of 12,836 cases across Kindergarten to 12th grade. The metric used is WIDA's overall composite score which is derived as follows: (reading scale score x 0.35) + (writing scale score x 0.35) + (listening scaled score x 0.15) + (speaking scaled score x 0.15). WIDA's overall composite score is weighted toward literacy skills (70%). This metric is used since a majority of WIDA member states use for AMAO calculations. WIDA uses the weighted composite score and create proficiency level scores which range from 1.0 to 6.0 in increments of 0.1-termed proficiency decimals. WIDA describes proficiency decimal scores as follows:

> The whole number indicates the student's language proficiency level as based on the WIDA ELP 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.[5]

Proficiency decimal scores are best treated as ordinal data; however, treating them as interval data, especially since composite scores are standardized by grade, should provide reasonably robust results.

Table 1 displays the average proficiency decimal gain by year in ELL program by cluster.

[5] WIDA Consortium (2008). *Assessing Comprehension and Communication in English State to State for English Language Learners-ACCESS for ELLs®: Interpretive Guide for Score Reports, Spring 2008*. Madison, WI: Author.

Table 1: Proficiency Decimal Growth by Years in Program by Cluster

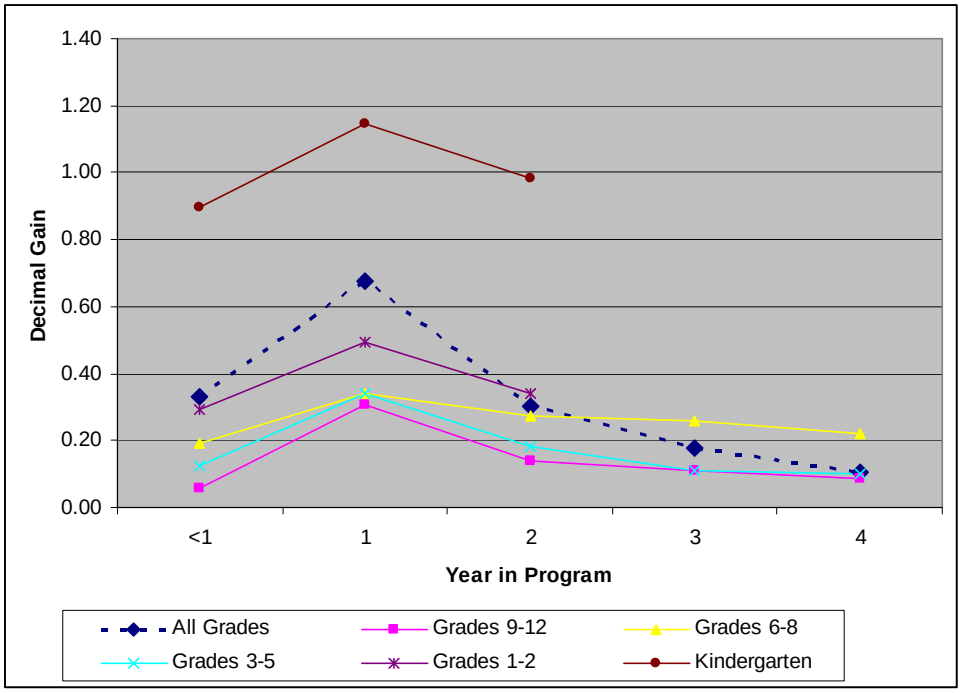
Length	Statistics	K	1	3	6	9	Total
<1	Mean	0.90	0.29	0.12	0.19	0.06	0.33
	N	232	232	269	198	148	1079
	SD	0.794	0.869	0.904	0.803	0.672	0.880
1	Mean	1.14	0.49	0.34	0.34	0.30	0.68
	N	1349	736	674	453	310	3522
	SD	0.735	0.849	0.753	0.660	0.631	0.834
2	Mean	0.98	0.34	0.18	0.27	0.14	0.30
	N	142	1221	679	402	292	2736
	SD	0.737	0.731	0.697	0.646	0.615	0.721
3	Mean	0.82	0.20	0.11	0.26	0.11	0.18
	N	6	742	474	283	168	1673
	SD	0.768	0.739	0.637	0.663	0.630	0.691
4	Mean		-0.05	0.10	0.22	0.09	0.11
	N		133	738	241	132	1244
	SD		0.705	0.641	0.632	0.576	0.643
5	Mean		0.15	0.10	0.23	0.15	0.14
	N		13	419	197	99	728
	SD		0.864	0.668	0.716	0.912	0.723
6	Mean		0.90	0.01	0.09	-0.09	0.03
	N		2	230	170	56	458
	SD		0.849	0.645	0.709	0.649	0.674
7	Mean			-0.16	0.02	-0.05	-0.03
	N			46	171	37	254
	SD			0.988	0.748	0.688	0.788

Let us say that we wish to use gain in proficiency level decimal scores as our metric for AMAO 1. The data from Table 1 can help us understand the relationships between time in program and grade level cluster as it relates to AMAO 1. In Table 1, the leftmost column lists years in ELL program, which spans from less than one year to 7 years. The mean (average), N-size and standard deviation (SD) are presented as statistics. Columns to the right of statistics column refer to Kindergarten (K), grade clusters 1-2 (1), 3-5 (3), 6-8 (6) and 9-12 (9). The far most column aggregates across grade clusters. Thus, students with less than one year in their ELL program (across grade clusters) gained on average 0.33 proficiency decimal points. In total, there were 1,079 students with less than one year in program with a standard deviation of 0.880. Notice that students with more years in ELL programs tend to have lower gain scores (with the exception of less than one year students). Notice also that as you increase grade cluster, the trend in gain scores also decreases regardless of the years in program. The trends just mentioned suggest that there is justification to create cohorts based on grade clusters. If this isn't done, districts with larger populations of higher grade students would be at a disadvantage—possibly just because they have higher grade students.

Figure 1 displays the same information in a line chart. Across grade based, with the exception of <1 year in program, we see a decreasing trend in gain across all grade level clusters. When interpreting both Table 1 and Figure 1, it is important to consider the variation in test scores. The standard deviations of scores are fairly large, suggesting that within grade clusters

there is a substantial amount of variation in how student grow. This variation could be attributed to several sources, e.g., program effects, language effects, within student effects, etc.

Figure 1: Graph of Average Proficiency Decimal Growth by Year in Program by Cluster



Thus far, creating cohorts based on grade level clusters and differentiating growth expectations by clusters would be permissible under this interpretation. As an example, let us take the 3-5 grade cluster and create an AMAO 1 expectation table (see Table 2).

Table 2: Example AMAO 1 Expectation Table for 3-5 Grade Cluster

Length in Program	Observed Decimal Gain	AMAO 1 Expected Gain
<1	0.12	0.15
1	0.34	0.35
2	0.18	0.25
3	0.11	0.15
4	0.10	0.15

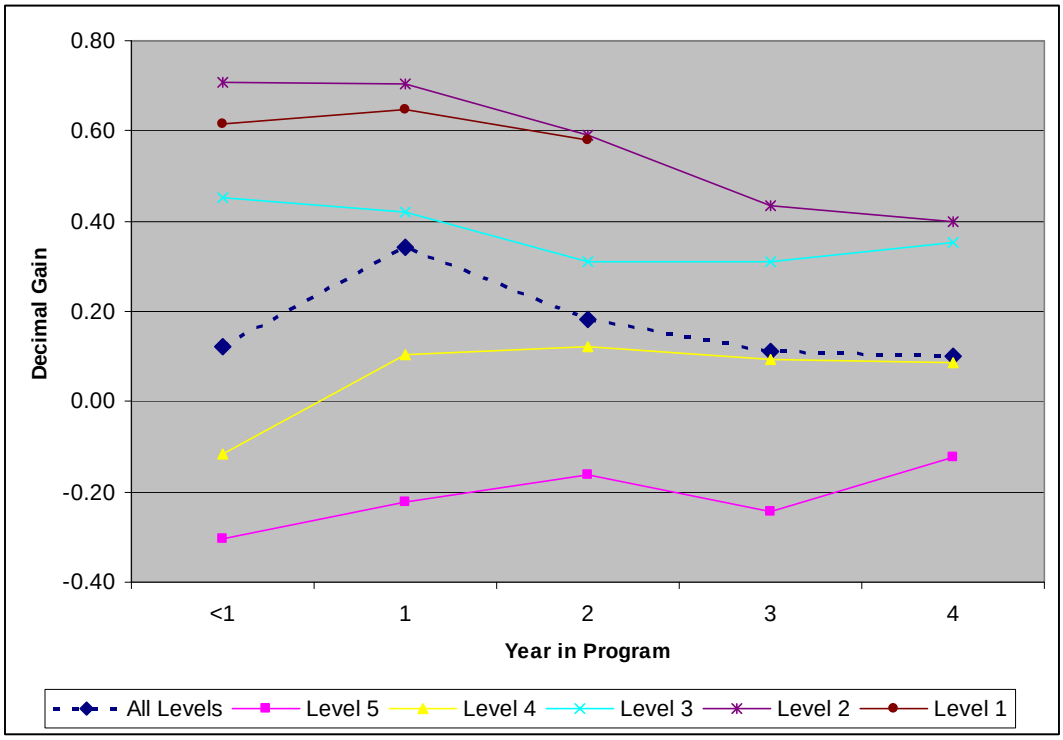
Let us say that we wish to set slightly higher AMAO 1 gain targets across program years, with a minimal gain expectations being 0.15 (at <1 year and greater than 2 years). Now let us take a closer look at the 3-5 grade cluster. Now we focus in not only on length in program but also the initial proficiency level by length in program. Table 3 displays this information.

Table 3: Grade Cluster 3-5 Decimal Gain by Year in Program by Proficiency Level

Length	Statistics	WIDA Initial Proficiency Level						Total
		1	2	3	4	5	6	
<1	Mean	0.62	0.71	0.45	-0.12	-0.31	-0.69	0.12
	N	41	30	55	71	51	21	269
	SD	0.830	0.783	0.642	0.753	0.811	1.128	0.904
1	Mean	0.65	0.70	0.42	0.10	-0.22	-0.38	0.34
	N	109	138	174	149	84	20	674
	SD	0.688	0.637	0.614	0.720	0.641	1.105	0.753
2	Mean	0.58	0.59	0.31	0.12	-0.16	-0.24	0.18
	N	24	105	185	187	136	42	679
	SD	0.684	0.624	0.658	0.702	0.633	0.338	0.697
3	Mean		0.43	0.31	0.09	-0.24	-0.26	0.11
	N		60	137	146	93	31	474
	SD		0.532	0.550	0.623	0.644	0.582	0.637
4	Mean		0.40	0.35	0.09	-0.12	-0.32	0.10
	N		59	182	229	192	65	738
	SD		0.626	0.580	0.563	0.604	0.531	0.641
5	Mean		0.40	0.30	0.08	-0.16	-0.16	0.10
	N		34	107	125	116	30	419
	SD		0.485	0.670	0.596	0.555	0.378	0.668
6	Mean		0.41	0.28	0.06	-0.41	-0.32	0.01
	N		18	69	65	55	21	230
	SD		0.805	0.503	0.580	0.619	0.446	0.645

Table 3 displays mean proficiency decimal gain by starting WIDA proficiency level. WIDA has 6 proficiency levels (1=Entering, 2=Beginning, 3=Developing, 4=Expanding, 5=Bridging, 6=Reaching). A majority of states set English language proficiency to be less than 6, which is why level 6 is shaded. What is immediately apparent is students at lower proficiency levels have substantially higher decimal gain scores across all program lengths. Said differently, students at lower proficiency levels tend to grow faster than students at higher proficiency levels. Cook, et al. coined the principal "lower is faster and higher is slower." This means that students at lower grades or proficiency levels tend to grow faster than their higher proficiency level/grade peers. If we take the average decimal gain of levels 1, 2 and 3 Table 3 and compare them to the expected gain shown in Table 2, we note that the average gain in these levels surpassed the expected AMAO 1 gain. Figure 2 graphically displays this relationship.

Figure 2: Graph of Cluster 3-5 Average Proficiency Decimal Gain by Year in Program



The average proficiency level gain for levels 4 and 5 are both under the "All Levels" gain line, which is lower than the expected growth shown in Table 2. What does this mean? If you have an influx of new students into a district who are at lower proficiency levels, they will tend to grow faster than students at higher levels. Thus, initial proficiency levels HAVE substantial influence on how students grow regardless of how long they've been in a language program. There are legitimate linguistic and maturational reasons for these differences and NOT accounting for them could (most likely will) misidentify LEAs that have large influxes of students at particular proficiency levels.

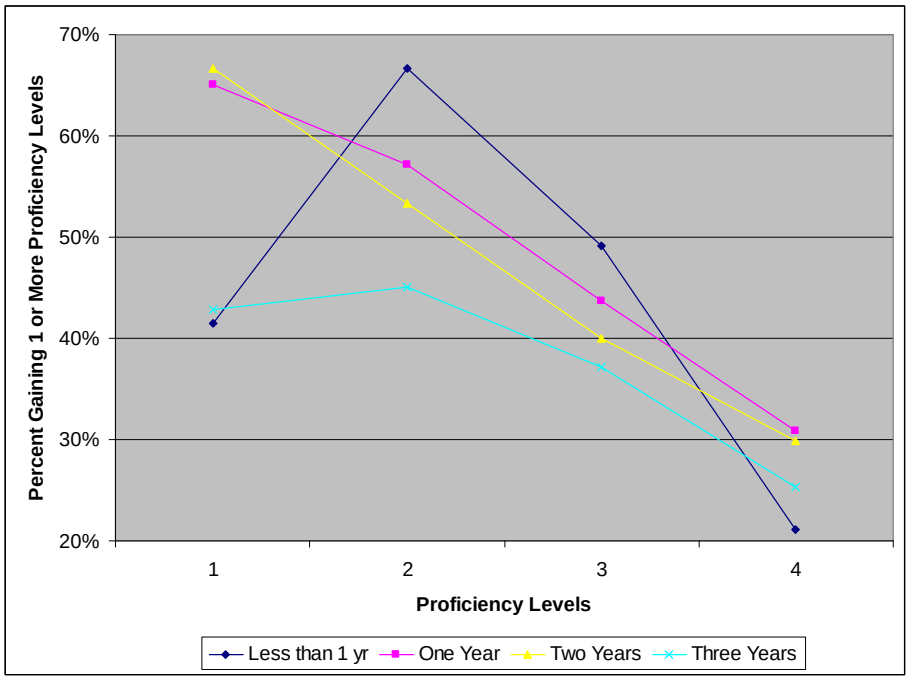
What we observe in Table 2 and Figure 3 may merely be a function of proficiency level decimal scores. Other states use the percent of students who gain one or more proficiency levels as the metric for AMAO 1. Let us look at the same grade level cluster with this in mind. Table 4 and Figure 3 display this analysis for the 3-5 grade cluster.

Table 4: Grade Cluster 3-5 Percent of Students Gaining One or More Proficiency Levels by Year in Program by Proficiency Level

Length in Program	Proficiency Levels	Grades 3-5	
		Count ≥1	% ≥1 Level or More
< 1	1	17	41.5%
	2	20	66.7%
	3	27	49.1%
	4	15	21.1%
	5	9	17.6%
	Total	88	32.7%
1	1	71	65.1%
	2	79	57.2%

Length in Program	Proficiency Levels	Grades 3-5	
		Count ≥1	% ≥1 Level or More
	3	76	43.7%
	4	46	30.9%
	5	14	16.7%
	Total	286	42.4%
2	1	16	66.7%
	2	56	53.3%
	3	74	40.0%
	4	56	29.9%
	5	28	20.6%
	Total	230	33.9%
3	1	3	42.9%
	2	27	45.0%
	3	51	37.2%
	4	37	25.3%
	5	16	17.2%
	Total	134	28.3%

Figure 3: Graph of Cluster 3-5 Percent of Student Gaining One or More Proficiency Levels by Length in Program and Initial Proficiency Level



The trend mentioned earlier with decimal scores is sustained with proficiency levels. In general, as proficiency levels increase the percent of students gaining levels decreases, in some cases quite substantially. The effect of proficiency levels on gain trends seen with the 3-5 grade cluster are similar to those in the Kindergarten, Grades 1-2, 6-8 and 9-12 clusters.

Initial proficiency levels make a difference on how student grow—regardless of how long they have been in a language program. The irony is the current interpretation, using only years in

program, would allow lower growth for LEAs with higher densities of low level students. Is that the intent? I don't believe so.

As seen in Tables 3 and 4, it is possible to array proficiency levels AND year in program together. Using these tables, AMAO 1 growth targets could be established accounting for differences in proficiency levels—having higher expectations for lower level/lower grade students. This would better address the spirit and letter of 3122(a)(2)(A). Failure to account for proficiency levels, in my view, ignores available evidence and is inconsistent the desire to implement scientifically researched policies, as so frequently mentioned in NCLB.

9. DETERMINING AMAOS FOR CONSORTIA

I have no comments on this interpretation.

10. IMPLEMENTATION OF CORRECTIVE ACTIONS UNDER TITLE III

I have no comments on this interpretation.