Technical Report # 1311

Analysis of Growth on State Tests for Students With Significant Cognitive Disabilities

Dan Farley

Jessica L. Saven

Gerald Tindal

Joseph F. T. Nese

University of Oregon

Published by

Behavioral Research and Teaching
University of Oregon • 175 Education
5262 University of Oregon • Eugene, OR 97403-5262
Phone: 541-346-3535 • Fax: 541-346-5689
http://brt.uoregon.edu

Note: This technical report was supported in part by the Oregon Department of Education (ODE) and the National Center on Assessment and Accountability in Special Education (NCAASE) grant # R324C110004 from the Institute of Education Sciences. The report does not reflect any endorsement by any of these organizations.

Abstract

Alternate assessments based on alternate achievement standards are designed to measure the academic achievement of students with the most significant cognitive disabilities. Because this population has not previously been included in large-scale testing programs, these assessments present unique measurement challenges. Probably the most significant issue is the inherent need for individualization in item presentation and response while maintaining rigorous levels of standardization. Additional measurement challenges are presented as states move toward implementation of growth models for accountability. In this report, we discuss four approaches to modeling the growth of students with significant cognitive disabilities. We apply several variations of transition matrix growth modeling for one state's alternate assessments and discuss the measurement challenges and policy considerations related to our findings.

Keywords: alternate assessment, students with significant cognitive disabilities, growth models, transition matrix

Overview

In this report we use a transition matrix to describe growth for students with the most significant disabilities who took Oregon's alternate assessments based on alternate achievement standards (AA-AAS), the Oregon Extended Assessments (ORExt), for statewide accountability purposes. Although the report is focused on Oregon's alternate assessment system, the challenges are common to the field in depicting change over time (growth) for students with the most significant cognitive disabilities.

Alternate assessments judged against alternate achievement standards (AA-AAS) are designed to measure the academic achievement of students with the "most significant cognitive disabilities" exclusively as part of the U.S. Department of Education's effort to ensure that "schools are held accountable for the educational progress of students with the most significant cognitive disabilities" (Title 1, 2003, p. 68698). Only students who are eligible for special education services through the Individuals with Disabilities Education Act (IDEA) are eligible to participate in the AA-AAS (NCLB, 2001; Title 1, 2003). The initial proposed rules from March 20, 2003 defined the terms "most significant cognitive disability" as having intellectual functioning and adaptive behavior that are three standard deviations or more below the mean (Title 1, 2003, p. 68700). However, the final regulations removed this strict definition to give states more flexibility and avoid placing "unwarranted reliance on an IQ score" (Title 1, 2003, p. 68704).

Though there is variation across states, the top three criteria used across most states are (a) the student has a significant cognitive disability, (b) eligibility decisions are made by Individualized Education Program (IEP) teams, and (c) substantial adjustments to curriculum are required in order to ensure access to the general education curriculum (Albus & Thurlow, 2012). Of the 58 states and territories that participated in this study, 40 did not allow disability label or characteristics to be used in participation decisions.

Nonetheless, sample statistics from six states indicate that the majority of students who participate in AA-AAS are eligible for special education services from the following three categories: intellectual disabilities, multiple disabilities, and autism (Kearns, Towles-Reeves, Kleinert, Kleinert, & Kleine-Kracht, 2011). These results are supported by a more recent survey conducted by the National Center and State Collaborative (NCSC) project (2012), with 56% of students participating in AA-AAS across 18 states having intellectual disabilities, 22%

with autism, 9% with multiple disabilities, 3% for other health impairment and "other," which may have included learning disabilities as well as developmental delay, according to the authors. No other category had more than 1% representation.

Students with significant cognitive disabilities (SWSCDs) can be difficult to assess in a standardized manner (Gong & Marion, 2006). Measurement challenges include trend analysis discrepancies, distribution assumptions, compounded standard errors, and multiple scales being used (Ho, 2008; Ho, 2009; Ho, Lewis, & MacGregor, 2009). Other issues appear with attempts to document growth including data system integrity, missing data, student mobility, student attrition, and scaling difficulties (Tindal, Schulte, Elliot, & Stevens, 2011). These measurement difficulties inherent within AA-AAS may also lead to different decisions made across tests when AA-AAS results are used in statewide accountability growth models that rely on adequate yearly progress (AYP).

Furthermore, as we discovered in conducting this study, additional challenges exist: (a) eligibility concerns, (b) participation (lack of a comparison group and grade retention), (c) variability in the performance levels selected, (d) within-group variability, and (e) reporting levels.

We organize the report in three sections. First, we define four different approaches to growth modeling. Second, we review a specific growth model that may be the most amenable to AA-AAS measurement challenges. Third, we address potential concerns and solutions in applying a transition matrix growth model to students participating in AA-AAS. We conclude by addressing some of the challenges noted above.

Growth Models

Growth models require multiple years of high-quality data for the same students. To show growth, a common scale also needs to be used over this time period. Given a data set that meets these basic criteria, at least four types of growth models can be considered in statewide accountability systems. In the presentation that follows, each model is defined in terms of benefits and limitations.

Status and Improvement (No Child Left Behind)

The first model we consider is the status and improvement model promulgated by No Child Left Behind (NCLB). This model provides a snapshot of academic performance at one point in time and compares the functioning, as defined by proficiency percentages, of successive groups (e.g., this year's fourth graders with last year's fourth graders). Results are reported at the school, district, and state levels. Accountability is defined by set percentages of students who meet proficiency standards that have been targeted over time (with the goal originally established for 100% of the population). Given this stringent standard, an exception is allowed (safe harbor) as long as schools are successful in reducing the percentage of students below proficient by 10% compared to the prior year.

Transition Matrix Model

In this model, student growth is depicted as changes in percentages of students at various performance standard levels (*Does Not Yet Meet, Nearly Meets, Meets, Exceeds* in Oregon) with the option to award points that add value to changes in these performance levels. For example, points can be awarded for students who perform at a higher level from one year to the next and subtracted for students who perform at a lower level from one year to the next, with no points for students who maintain performance levels from one year to the next. This approach allows for scores from tests on different scales to be aggregated on a common scale of percentage change in levels. Alternate assessments are rarely scaled across grades. Transition matrices can depict growth nonetheless.

Residual Gain and Value Added Model

The Residual Gain and Value Added Model (ResVAM) conditions current performance by past performance in calculating residuals between the predicted score and actual score in the current year. Residual scores near zero denote growth consistent with predicted scores, positive values reflect growth that exceeds prediction, and negative scores reflect growth that is less than predicted. Variations of this model use multiple years of prior performance or other factors, such as student background characteristics to condition current performance.

Multilevel Growth Models

Multilevel Growth Models (MGM) fit growth trajectories of each student over time with both starting level and slope analyzed. Further levels can be used to condition growth, including student characteristics (e.g., demographics such as gender, race-ethnicity, English language learner status, or special education services received) or teacher and school contexts (e.g., class size, years of experience, building level programs). MGMs are unique in that variance of these nested levels is appropriately partitioned and therefore are more accurate than difference or residual score models (Koretz & Hamilton, 2006).

AA-AAS Growth Model Applications

The status and improvement model (S-I) is designed to answer the question: What percentage of this year's students met AYP? The transition matrix model (TM) answers the question: Are students making adequate progress across performance levels? For the residual gain score model (R-G), the question is: How much residual gain was produced by a group? Finally, the essential question for the multi-level growth model (MGM): What is the starting level and slope of growth for students (and as conditioned further over aggregations of teacher and school)? In Table 1, provided by the National Center on Assessment and Accountability in Special Education (NCAASE), growth models are compared on a number of features.

A review of Table 1 demonstrates a primary advantage of TM approaches to modeling growth for SWSCDs. TM approaches allow for the inclusion of different tests with different scales. It also allows for across category and within category growth projections. For these and other related reasons, the TM approach will be the approach to modeling growth for AA-AAS addressed in this technical report.

Table 1

The requirements and capabilities of the four growth models

Data Requirements	**S-I**	**TM**	**R-G**	**MGM**
Database of matched student records over time (student ID)	No	Yes	Yes	Yes
Common scale	No	No	Yes	Yes
Precision and accuracy evaluated	Yes	Yes	Yes	Yes
Confidence interval	Indiv. Grps.	Std. Errors	Error Var.	Error Var.
Includes students with missing scores	Yes	No	No	Yes
Affected by cohort stability	Yes	Yes	Yes	Yes
Handles non-linear growth	No	No	No	Yes
Includes results from alternate tests (different scales)	No	Yes	No	No
Student performance standards in definition of growth	Yes	Yes	No	No

Note: This table used with permission from the principal investigators of NCAASE

Transition Matrix Growth Model

Research Questions

In the remaining two sections, we first present data using the transition matrix growth model and then address two critical questions that challenge states when analyzing growth for SWSCD: What models are feasible and what are the measurement challenges?

Methods

We include data from two successive years for a sample of students in grades 3-8. As can be seen in Table 2, a sizeable group of students could not be included in growth calculations because they were either 3rd or 8th graders in the second year of the transition year pairs. Third graders who entered Cohort 1 in 2009 could not be included in grade 3 growth calculations as no prior year test exists (the statewide assessment begins in grade 3). Eighth graders who entered Cohort 1 in 2008 could not be included in growth calculations because no subsequent year test exists (there is no grade 9 test). These results generalize to Cohort 2. Otherwise, a very significant number of students are missing scores for unknown reasons with a small number missing because they were retained. Because there is no prior or subsequent year test, it was not possible to include grade 11 growth calculations.

Table 2

Missing Data from Successive Years (Cohorts)

Reason for Count	**Cohort 1 (2008/09)**	**Cohort 2 (2010/11)**
Beginning Total	6,722	7,181
3rd Graders Missing Comparison Group		
- *Spring 2009 for Cohort 1*	1,116	1,217
- *Spring 2011 for Cohort 2*		
8th Graders Missing Comparison Group		
- *Spring 2008 for Cohort 1*	490	508
- *Spring 2010 for Cohort 2*		
Missing a year	2,183	1,986
Retained	59	40
Total	**2,874**	**3,430**

Population

All 11 Oregon school-age disability categories are represented in both cohorts. In 2008/09, the primary disability categories were Intellectual Disability – ID (26%), Specific Learning Disability - SLD (24%), Autism Spectrum Disorder - ASD (16%), Other Health Impairments - OHI (7%), and Communication Disorder - CD (10%). In 2010/11, the primary category memberships were ID (28%), ASD (19%), SLD (18%), CD (11%) and OHI (10%). Some students switched categories between years, a total of 217 students in the 2008/09 transition (6%), and 247 students in the 2010/11 transition (7%). Tables 3 & 4 below depict disability categories where more than 10 students shifted in a given transition period.

Table 3
Disability category shifts for the 2008/09 transition years

	Disability Category 2009					
Disability Category 2008	**ID**	**CD**	**OHI**	**Autism**	**SLD**	**Total**
Intellectual Disability (ID)	0	2	7	5	2	**16**
Communication Disorder (CD)	23	0	10	5	37	**75**
Other Health Impairments (OHI)	22	6	0	7	4	**39**
Autism Spectrum Disorder	7	2	1	0	1	**11**
Specific Learning Disability (SLD)	11	7	3	1	0	**22**
Total	**63**	**17**	**21**	**18**	**44**	**163**

*54 other students shifted disability categories in other categories (<10/category).

Table 4
Disability category shifts for the 2009/10 transition years

	Disability Category 2011					
Disability Category 2010	**ID**	**CD**	**OHI**	**Autism**	**SLD**	**Total**
Intellectual Disability (ID)	0	9	6	5	2	**22**
Communication Disorder (CD)	29	9	8	4	47	**97**
Other Health Impairments (OHI)	18	2	0	5	2	**27**
Autism Spectrum Disorder	10	0	4	0	3	**17**
Specific Learning Disability (SLD)	12	16	7	1	0	**36**
Total	**69**	**36**	**25**	**15**	**54**	**199**

*48 additional students shifted disability categories in other categories (<10/category).

Four categories may be classified as requiring less support in that intensive instruction is required primarily in academic or language systems rather than in both academic and physical-social environments. With the 2008/09 transition, 49.69% of students shifted from a disability requiring less support (i.e., SLD, CD, OHI) to disabilities requiring more support (i.e., ID, ASD). Shifts from ID and ASD to categories that typically require less support also occurred (12.26%). Some students continued to receive more support, but switched from ASD to ID or vice versa (7.36%). In the 2010/11 transition, 42.21% of students shifted from a disability requiring less support to disabilities requiring more support. Shifts from ID and ASD to categories that typically require less support also occurred (14.57%). Some students continued to receive more support, but switched from ASD to ID or vice versa (7.54%).

Analyses

We constructed cross tabulation tables comparing the frequencies of each of the four performance categories in order to generate transition matrices for both the spring 2008 transition to spring 2009 (2008/09) and the spring 2010 transition to spring 2011 (2010/11). The four categories in each matrix include: *Does Not Yet Meet* (DNYM), *Nearly Meets* (NM), *Meets* (M), and *Exceeds* (E). Students earned a point for moving up one performance level and lost a point for moving down one performance level. For example, a student who moved up one performance level from DNYM to NM generated a +1. A student who fell a performance level from E to M generated a -1.

*Transition analysis 1***.** Tables 4-8 below depict the 2008/09 transition matrix analysis for reading in grades 4-8. The cells contain the number of students who performed at each level from spring 2008 to spring 2009. For example, in grade 4 in the *Does Not Yet Meet* (DNYM) category, 142 students who performed at DNYM in 2008 performed at the same level in 2009; 20 students moved from the DNYM category in 2008 into the *Nearly Meets* (NM) category in 2009; eight students moved from the DNYM category in 2008 to the *Meets* (M) category in 2009; and, three students moved from the DNYM category in 2008 all the way up to the *Exceeds* (E) category in 2009. Alternatively, 122 students who performed at E in 2008 matched their performance in 2009; 23 students who performed at E in 2008 dropped to M in 2009; two students who performed at E in 2008 dropped to NM in 2009; and, no students dropped from E to DNYM.

Adequate Yearly Progress results were calculated by comparing the 2008 performance level to the 2009 performance level. Students received +1 points for each performance level increase, or received a -1 for each performance level decrease. Students who performed at the same level received a 0 points for growth, with the exception of those who maintained Exceed status, who received +1 points. We call this model the AYP+1 model. Because growth is not possible at the Exceed level, we awarded a bonus point for maintenance in this category. For example, a student who achieves at the highest level and falls to the lowest level receives a -3 AYP+1 rating, as the student fell three performance levels that year. A student who rose from DNYM to E would receive a +3 rating for having risen three levels, etc.

If a bonus point is not included for students maintaining Exceeds, the AYP ratings, decreases in the 2008/09 transition from an average AYP rating of 113.8 to an average AYP+1 rating of -41.2.An alternative to the AYP+1 rating system is the AYP +2 rating system, which awards students +2/-2 points for increasing/decreasing each performance level, 1 point for maintaining at the *Exceeds* level, and 0 points for maintaining in all other categories. The AYP+2 approach is an attempt to counter the effect of weighting maintenance at the *Exceeds* level (and therefore should not be awarded as many points as students who improve levels). Both models are displayed in Tables 5-9 below.

Table 5
AYP ratings for Grade 4

GRADE 4		**2009**			**AYP +1**	**AYP +2**
2008	Does Not Yet Meet	Nearly Meets	Meets	Exceeds*		
Does Not Yet Meet	138	20	8	3	45	90
Nearly Meets	33	50	31	6	10	20
Meets	14	41	131	109	40	80
Exceeds	0	2	22	120	94	68
Totals	**185**	**113**	**192**	**238**	**189**	**258**

Table 6
AYP ratings for Grade 5

GRADE 5		**2009**			**AYP +1**	**AYP +2**
2008	Does Not Yet Meet	Nearly Meets	Meets	Exceeds*		
Does Not Yet Meet	165	15	5	1	28	56
Nearly Meets	35	45	34	6	11	22
Meets	13	23	142	76	27	54
Exceeds	5	1	26	107	64	21
Totals	**218**	**84**	**207**	**190**	**130**	**153**

Table 7
AYP ratings for Grade 6

GRADE 6	2009				AYP +1	AYP +2
2008	Does Not Yet Meet	Nearly Meets	Meets	Exceeds*		
Does Not Yet Meet	129	44	21	1	89	178
Nearly Meets	11	28	62	3	57	114
Meets	5	11	79	64	43	86
Exceeds	5	1	31	90	42	-6
Totals	**150**	**84**	**193**	**158**	**231**	**372**

Table 8
AYP ratings for Grade 7

GRADE 7	2009				AYP +1	AYP +2
2008	Does Not Yet Meet	Nearly Meets	Meets	Exceeds*		
Does Not Yet Meet	99	28	4	0	36	72
Nearly Meets	16	34	8	1	-6	-12
Meets	11	26	87	53	5	10
Exceeds	3	0	23	64	32	0
Totals	**129**	**88**	**122**	**118**	**67**	**70**

Table 9
AYP ratings for Grade 8

GRADE 8	2009				AYP +1	AYP +2
2008	Does Not Yet Meet	Nearly Meets	Meets	Exceeds*		
Does Not Yet Meet	114	7	0	0	7	14
Nearly Meets	33	38	7	2	-22	-44
Meets	8	30	51	24	-22	-44
Exceeds	3	4	39	45	-11	-67
Totals	**158**	**79**	**97**	**71**	**-48**	**-141**

*Note: Tables 4-8 * The scores in this column include a bonus point for maintenance at the Exceeds level; removing this bonus point is significant, as it reduces the overall AYP rating by the number in the Exceeds column.*

Results. The results displayed for the 2008/09 transition years demonstrate several trends. First, the majority of students perform at the same level from one year to the next (61%). Second, it becomes more difficult to maintain positive AYP ratings as students proceed through the grade levels. With the exception of the 6th grade, the results reflect a downward trend, ultimately resulting in negative growth scores in 8th grade.

Table 10 below conveys the overall shifts in terms of performance categories from a numerical perspective, providing a different way of viewing the data presented above, where each grade level is displayed on one row. These data are based on the same students from the previous tables. Note that a small number of students advanced multiple grade levels, which presents an interesting problem for future analyses.

Table 10
Overall performance category shifts for the 2008/09 transition years

	Change in Performance Category							
Grade	-3	-2	-1	0	1	2	3	**Total**
4	0	16	96	439	160	14	3	**728**
5	5	14	84	459	125	11	1	**699**
6	5	6	53	326	170	24	1	**585**
7	3	11	65	284	89	5	0	**457**
8	3	12	102	248	38	2	0	**405**
Total	**16**	**59**	**400**	**1756**	**582**	**56**	**5**	**2874**

The majority of students maintained their performance level (1756/2874 = 61%). A total of 475 students dropped one or more performance levels (475/2874 = 16.5%), while 643 students advanced one or more performance levels (643/2874 = 22%). The trend is upward.

In Table 11, a different set of categories has been created using ranges of change in the scaled score. With this system, it is possible to use more discrete (or selective) representations of student growth, which in turn may be more sensitive to increases or decreases in growth.

Table 11
Performance category shifts based on scaled score ranges to create a seven-categories

	Change in RIT Score							
	Decrease			Static	Increase			
Grade	> 60 pt ⬇	31 to 60 pt ⬇	1 to 30 pt ⬇	0	1 to 30 pt ⇧	31 to 60 pt ⇧	> 60 pt ⇧	**Total**
4	0	10	187	29	498	4	0	**728**
5	0	12	199	32	447	9	0	**699**
6	4	26	340	36	176	2	1	**585**
7	0	5	183	20	238	11	0	**457**
8	2	6	172	23	198	4	0	**405**
Total	**6**	**59**	**1081**	**140**	**1557**	**30**	**1**	**2874**
(%)	(.2)	(2)	(37.6)	(4.9)	(54.18)	(1)	(.03)	

In Table 12, we provide an example of the same growth model used above but with the two AYP calculations and the seven RIT-score ranges (30 points each, except for the maintenance level). As before, students earned 1 point for moving up a level and lost one point for moving down a level. The 2-point model is also presented.

Some interesting patterns are worthy of note. The model is much more sensitive to change of performance, as the categories are more discrete. The majority of students are indeed growing and not remaining static, as found with the four-level model. The general decrease in growth from grade 3 to 8 is not as obvious as it was with the four-level analysis. Overall, the trend is upward, with more students increasing (1,588) compared to decreasing (1,146). Significant grade level differences are apparent. In fact, 6th grade went from the highest performing grade in the four-level approach to the lowest performing grade in the seven-level approach simply due to the number of performance levels included in the calculations.

Table 12
Demonstrating the seven-level category analysis

Grade	AYP +1	AYP +2
4	299	598
5	242	484
6	-223	-447
7	67	134
8	16	32
Total	**401**	**801**

*Transition analysis 2***.** Tables 13-17 below depict the 2010/11 transition matrix analysis for reading in grades 4-8 using the same calculations as the Transition Analysis 1. For example, in grade four in the *Does Not Yet Meet* (DNYM) category, 132 students who performed at DNYM in 2010 also performed at the same level in 2011; 35 students moved from the DNYM category in 2010 into the *Nearly Meets* (NM) category in 2011; 10 students moved from the DNYM category in 2010 to the *Meets* (M) category in 2011; and, one student moved from the DNYM category in 2010 all the way up to the *Exceeds* (E) category in 2011. At the same time, 138 students who performed at Exceed in 2010 matched their performance in 2011; 35 students who performed at Exceed in 2010 dropped to Meets in 2011; two students who performed at level Exceed in 2010 dropped to Not Meeting in 2011; and, one student dropped from Exceed in 2010 all the way down to Does Not Yet Meet in 2011.

Table 13
AYP ratings for Grade 4

GRADE 4			**2011**		**AYP +1**	**AYP +2**
2010	Does Not Yet Meet	Nearly Meets	Meets	Exceeds*		
Does Not Yet Meet	127	35	10	1	58	116
Nearly Meets	35	54	41	6	18	36
Meets	15	50	163	95	15	30
Exceeds	1	2	34	138	97	56
Totals	**178**	**141**	**248**	**240**	**188**	**238**

Table 14
AYP ratings for Grade5

GRADE 5			**2011**		**AYP +1**	**AYP +2**
2010	Does Not Yet Meet	Nearly Meets	Meets	Exceeds*		
Does Not Yet Meet	170	18	3	0	24	48
Nearly Meets	51	62	40	6	1	2
Meets	17	31	146	79	14	28
Exceeds	4	2	59	158	83	8
Totals	**242**	**113**	**248**	**243**	**122**	**86**

Table 15
AYP ratings for Grade 6

GRADE 6	2011				AYP +1	AYP +2
2010	Does Not Yet Meet	Nearly Meets	Meets	Exceeds*		
Does Not Yet Meet	133	45	31	1	110	220
Nearly Meets	6	22	45	8	55	110
Meets	3	15	118	53	32	64
Exceeds	0	1	51	105	52	-1
Totals	**142**	**83**	**245**	**167**	**249**	**393**

Table 16
AYP ratings for Grade 7

GRADE 7	2011				AYP +1	AYP +2
2010	Does Not Yet Meet	Nearly Meets	Meets	Exceeds*		
Does Not Yet Meet	138	25	4	0	33	66
Nearly Meets	17	42	16	4	7	14
Meets	7	28	122	58	16	32
Exceeds	2	0	22	106	78	50
Totals	**164**	**95**	**164**	**168**	**134**	**162**

Table 17
AYP ratings for Grade 8

GRADE 8		**2011**			**AYP +1**	**AYP +2**
2010	Does Not Yet Meet	Nearly Meets	Meets	Exceeds*		
Does Not Yet Meet	142	7	3	1	16	32
Nearly Meets	30	67	8	0	-22	-44
Meets	1	46	74	50	2	4
Exceeds	1	2	23	94	64	34
Totals	**174**	**122**	**108**	**145**	**60**	**26**

*Note: Tables 12-16 * The scores in this column include a bonus point for maintenance at the Exceeds level; removing this bonus point is significant, as it reduces the overall AYP rating by the number in the Exceeds column.*

Comparison of transition matrices 1 and 2. The results displayed for the 2010/11 transition years establish several patterns similar to the 2008/09 transition. The majority of students performed at the same level from one year to the next (64%). Additionally, the trend continued with lower AYP ratings in higher grades. With the exception of the 6th grade, the overall results reflect a downward trend, with minimal growth scores for 8th grade.

In Table 18 the shifts in performance categories are displayed using the same restrictions as the 2008/09 analyses. As in 2008/09, a small number of students advanced multiple grade levels.

Table 18
Overall performance category shifts for the 2010/11 transition years

	Change in Performance Category							
Grade	-3	-2	-1	0	1	2	3	**Total**
4	1	17	119	482	171	16	1	807
5	4	19	141	536	137	9	0	846
6	0	4	72	378	143	39	1	637
7	2	7	67	408	99	8	0	591
8	1	3	99	377	65	3	1	549
Total	**8**	**50**	**498**	**2181**	**615**	**75**	**3**	**3430**

The majority of students maintained their performance level (2181/3430 = 64%). A total of 556 students dropped one or more performance levels (556/3430 = 16%), while 693 students advanced one or more performance levels (693/3430 = 20%). As with the 2008/09 transition, the trend for improvement increases.

Results from the 2010/11 transition are presented below in Table 19, using scale based categories to both increase the number of categories or more selectively target the ranges of them. In the end, the outcomes are similar to those we found in the 2008/09 transition.

Table 19
Performance category shifts based on a seven-category approach

	Change in RIT Score							
		Decrease		Static		Increase		
Grade	> 60 pt ↓	31 to 60 pt ↓	1 to 30 pt ↓	0	1 to 30 pt ⇧	31 to 60 pt ⇧	> 60 pt ⇧	**Total**
4	0	6	158	51	570	22	0	**807**
5	0	15	216	41	557	17	0	**846**
6	1	14	381	49	182	10	0	**637**
7	3	6	136	40	396	10	0	**591**
8	1	5	147	27	358	10	1	**549**
Total	**5**	**46**	**1038**	**208**	**2063**	**69**	**1**	**3430**
(%)	(.1)	(1.3)	(30.3)	(6.1)	(60.15)	(2.0)	(0.0)	

We see similar results as the 2008/09 transition, but even greater gains in the positive 1 to 30 point column. Again, for students in 6th grade, the number who decrease categories exceed the number who increase categories..

An example of applying the same growth model procedures used above at the performance level is elaborated below in Table 20 using RIT-score ranges of 30 points. The results replicate the trends of the 2008/09 transition matrix.

Table 20
Demonstrating the seven-level category analysis

Grade	AYP Rating	AYP +2
4	444	888
5	345	690
6	-210	-420
7	259	518
8	219	437
Total	**1057**	**2113**

Discussion

Transition matrices allow states to track the change of students in proficiency categories, which is important given the contingencies of NCLB with 100% of the population expected to become proficient. They also, however, present some challenges that states need to address if they are to be used in any accountability system. In the remainder of the technical report, we address these challenges and then conclude by reflecting on the limitations of our study.

Challenges

*Transition matrix approach relies on NCLB status-based approaches***.** The concerns highlighted by NCAASE researchers (2011) are systems-level issues that apply to all states attempting to implement growth models. These issues include data system integrity, missing data, student mobility, student attrition, and scaling. Other challenges also appear and are reflected in our study: eligibility concerns, participation (lack of comparison groups and other factors), grade retention, number of performance levels selected, homogeneity (within-group variability), and reporting levels. Furthermore, a number of other measurement challenges from NCLB status-based models also appear and are not addressed in this study: standard setting procedures, cut score choices across tests, trend analysis discrepancies, distribution assumptions, compounded standard errors, and multiple scales (Ho, 2008; Ho, 2009; Ho, Lewis, & MacGregor, 2009).

Eligibility concerns. Although alternate assessments are designed for students with the most significant cognitive disabilities, the population of students participating in the AA-AAS is extremely varied and includes those from every disability group. Some would argue that students with specific learning disabilities do not belong in the population of the students eligible to take the AA-AAS.

This issue actually may be explained by the criteria used for recommending this test option adopted in most states: (a) a significant cognitive disability exists, (b) modified instruction is required, (c) extensive support for skill generalization is needed, (d) modified curriculum are needed, and (e) the disability category is not the basis for eligibility (Cameto, et al., 2009). Oregon's criteria do not include the expectations that all eligible students (a) require modified instruction and curriculum (meaning significant reductions to the general education curriculum in order to access the content), and (b) need extensive support to generalize skills across all contexts (not limited to school content areas). Adaptive behavior deficits, which are concomitant with significant cognitive disabilities, also are not addressed by Oregon's current criteria.

Yet, as we determined in this study, students from every disability category participated in the AA-AAS option. Given Oregon's significant percentages of students who have specific learning disabilities (26-28%) and communication disorders (10-11%), populations for whom the test is really not designed to serve, eligibility criteria may need to be refined (see *Appendix A*) to determine whether or not the appropriate group of students is being targeted for AA-AAS participation. There were also a small number of students who participated in the ORExt who had no recorded disability code.

Participation. Lack of a comparison group is certainly a challenge when using a transition matrix to document growth. Our dataset had 1,116 students enter as 3rd graders in 2009; 1,217 students entered the dataset as 3rd graders in 2011. These students could not be included in growth analyses, as there was no comparison group (earlier performance as 2nd graders). Similar concerns are noted for 8th graders who entered in the first year of each cohort. Eleventh graders were not possible to include. States must determine how to include these three grade levels in accountability reporting, or the legislative requirements surrounding accountability must be adapted for growth models. This challenge actually generalizes across all growth models and is not specific to transition matrices.

Many students had missing scores in one year or the other. In our data set, over half of the students could not be included in analyses because they did not participate in both spring test administrations (2008/09 and/or 2010/11). Students who participate in one year of testing, but not the following year, were excluded (n = 2,183 in 2008/09; n = 1,986 in 2010/11). A more subtle issue in interpreting growth is the absence or presence of the student not in the data set but in the district (for which AYP is applied). Many students were not in the same school from one year to the next. These students may be included in district- or state-level AYP reporting, but will likely present challenges at the school level and in particular in inferring the meaning of growth (or lack thereof).

Retention. This issue is also present regardless of the model for analyzing growth. In our sample, some students were retained in each set of transition years (n = 59 in 2008/09; n = 40 in 2010/11). From spring 2008 to spring 2009, 14 students were retained in 5th grade, receiving instruction at a 5th grade level, yet were tested in 2009 using a middle school test form (we know this because their score is labeled as “valid” in the ODE data set). Not only is the assignment of the test level an issue, but the inference of proficiency becomes problematic. The instructional level that these students were exposed to during the school year may not necessarily have prepared them for the rigor of the assessment that they took in the spring. Of the 40 students retained from spring 2010 to spring 2011, nine stayed 5th graders, also changing the test form they took in 2011.

Homogeneity. A critical assumption in making growth comparisons is that the cohort population is the same from one year to the next. We would like to assume that the cohort, once it has been identified and filtered, can reasonably be considered homogeneous because all growth is on the same population of students. However, students shift in their disability categories from one year to the next so any more fine-grained analyses of growth by disability may be confounded. In most cases, students shifted from a disability requiring less intensive supports such as SLD or OHI, to the ID category, which typically requires more intensive supports. In other cases, students moved from a more intensive support category to a less intensive category. This potentially calls eligibility decisions into question.

Neither of these issues is unique to Oregon. Within-group heterogeneity of student disability categories is relatively well-established notion in the field, particularly with regard to academic achievement in reading and mathematics (Blackorby, Chorost, Garza, & Guzman, 2005; Blackorby & Wagner, 2005; Wei, Blackorby, & Schiller, 2011; Wei, Lenz, & Blackorby, 2012).

Reporting levels. Several challenges exist in using growth models for AA-AAS, in and beyond Oregon. Our findings are based on one state’s data for two sets of consecutive years. The sample sizes are sufficient for making AYP determinations at the macro level. However, the challenge of providing schools, and possibly some smaller districts, with an AYP rating when they may have very few, if any, students taking AA-AAS is an important planning consideration for the field.

In addition, we must determine how to treat the highest performance level (e.g., in Oregon's case, the *Exceeds* category). In our study, we awarded a bonus point to students who maintained performance at the *Exceeds* level, as it is not possible for them to improve categorically. If students in the *Exceeds* category are not given points for maintenance, the growth picture is impacted severely as it is difficult to make AYP as students advance through the grades; this challenge is likely to be compounded without some type of bonus system for the *Exceeds* level.

It is perhaps more concerning that different scales produce different results when implementing a TM model (e.g, the number of performance levels used affects the outcomes). We arrived at markedly different results when we used the existing four-level proficiency analyses than we did when we used a RIT-score range (of 30 points).

The TM model also introduces new challenges related to determining how much growth is sufficient at the student, school, district, and state levels. The state's existing standard setting procedures, including cut scores and proficiency level descriptors, can be used within the TM model. However, the model produces holistic AYP ratings, which will need to be analyzed for sufficiency.

Limitations and Future Directions

Limitations. This report presents one growth model, a specific variation of the Transition Matrix approach, for states to consider for AA-AAS. Clearly, our results are limited to reading in the grades tested. The results are exploratory in nature and should not be generalized. Yet, the measurement challenges are very likely to generalize across states that are implementing growth models.

Future directions. The TM approach is very flexible and feasible to implement with existing status-based performance structures. While this model is efficient and appears to hold some promise, it is shared not as a standard for the AA-AAS field to adopt, but as an objectification of how many measurement challenges the field faces in implementing growth models in a robust manner for SWSCDs. With AA-AAS, only a limited range of possibilities can be investigated.

In the end, the field needs to define at the school, district, and state levels how much growth is enough. It is hoped that future federal and state policies address the needs surrounding growth models to support their implementation, as the move from status-based models toward growth models is progressing.

It appears unlikely that states will be in a position to implement valid growth models, even Transition Matrices, for SWSCDs until they have:

- improved standard setting procedures or replaced these procedures with a relevant statistical methodology,

- developed statistical scaling and distribution correction techniques that allow for cross-test comparisons,
- developed, maintained, and increased data system integrity,
- accounted for attrition/missing values in a justifiable manner,
- accounted for grade level and disability category fluctuations/eligibility considerations
- determined a manner in which the lack of a comparison group can be addressed,
- defined how much growth is sufficient (particularly at the school level),
- ensured that the growth model approach selected is consistent with the state's overall conceptual and practical assessment model, and
- ensured that a valid system is constructed, with emphases upon consistency between and among different growth model methodologies.

Admittedly, these concerns may disappear if states move forward as a Tier II state with either the National Center and State Collaborative (NCSC) or the Dynamic Learning Maps (DLM) consortia. Both promise to deliver an entire AA-AAS assessment (formative and summative)-curriculum-professional development system to states (http://www.ncscpartners.org & http://dynamiclearningmaps.org).

References

Albus, D., & Thurlow, M. (2012). Alternate assessments based on alternate achievement standards (AA-AAS) participation policies. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Betebenner, D.W. (2008). Norm- and criterion-comparisond student growth. Retrieved from http://www.nciea.org/publications/normative_criterion_growth_DB08.pdf

Blackorby, J., Chorost, M., Garza, N., & Guzman, A. (2005). The academic performance of elementary and middle school students with disabilities. *SEELS Engagement, academics, social adjustment, and independence: The achievements of elementary and middle school students with disabilities.* Retrieved from http://www.seels.net/designdocs/engagement/All_SEELS_outcomes_10-04-05.pdf

Blackorby, J., & Wagner, M. (2005). Students with disabilities in elementary and middle school: Progress among challenges. *SEELS Engagement, academics, social adjustment, and independence: The achievements of elementary and middle school students with disabilities*. Retrieved from http://www.seels.net/designdocs/engagement/All_SEELS_outcomes_10-04-05.pdf

Cameto, R.,Knokey, A.-M., Nagle, K., Sanford, C., Blackorby, J., Sinclair, B., et al. (2009). *National profile on alternate assessments based on alternate achievement standards. A report from the national study on alternate assessments* (NCSER 2009-3014). Menlo Park, CA: SRI International.

Gong, B., & Marion, S. (2006). *Dealing with flexibility in assessments for students with significant cognitive disabilities* (Synthesis Report 60). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Hambleton, R. (2001). Setting performance standards on educational assessments and criteria for evaluating the process. In G.J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives*. (pp. 89-116). Mahwah, NJ: Lawrence Erlbaum Associates.

Ho, A.D., Lewis, D.M., & MacGregor Farris, J.L. (2009). The dependence of growth-model results on proficiency cut scores. *Educational Measurement: Issues and Practice, 28* (4), 15-26.

Ho, A.D., (2009). A nonparametric framework for comparing trends and gaps across tests. *Journal of Educational and Behavioral Statistics*, *34* (2), 201-228.

Ho, A.D., (2008). The problem with "proficiency": Limitations of statistics and policy under No Child Left Behind. Educational Researcher, *37* (6), 351-360.

Kearns, J., Towles-Reeves, E., Kleinert, H., Kleinert, J. O., & Kleine-Kracht, M. (2011). Characteristics of and implications for students participating in alternate assessments based on alternate academic achievement standards. *The Journal of Special Education, 45*, 3-14. doi: 10.1177/0022466909344223

Koretz, D. M., & Hamilton, L. S. (2006). Testing for accountability in K-12. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 531-578). Westport, CT: Praeger.

Linn, R. L., (2003). Performance standards: Utility for different uses of assessments. *Education Policy Analysis Archives*, 11 (31). Retrieved October 11, 2011 from http://epaa.asu.edu/epaa/v11n31/.

National Center and State Collaborative. (2010). National Center and State Collaborative General Supervision Enhancement Grant (NCSC GSEG): U.S. Department of Education.

National Resource Center on Assessment and Accountability for Special Education (2013). http://www.ncaase.com/docs/NarrativeV15_NationalRDCtrFINAL91410v4.pdf

The No Child Left Behind Act (NCLB), Pub. L. No. 107-110 (2001).

Towles-Reeves, E., Kearns, J., Flowers, C., Hart, L., Kerbel, A., Kleinert, H., Quenemoen, R. & Thurlow, M. (2012). *Learner characteristics inventory project report (A product of the SCSC validity evaluation)*. Minneapolis, MN: University of Minnesota, National Center and State Collaborative.

Title 1 – Improving the academic achievement of the disadvantaged; Final rule. 34 CFR.200. (2003).

Wei, X., Blackorby, J., & Schiller, E. (2011). Growth in reading achievement in a national sample of students with disabilities ages 7 to 17. *Exceptional Children, 78*, 89-106.

Wei, X., Lenz, K. B., & Blackorby, J. (2012). Math growth trajectories of students with diabilities: disability category, geneder, racial, and socioeconomic status differences from ages 7 to 17. *Remedial and Special Education*, published online 16 July 2012 doi: 10.1177/0741932512448253