Student Growth and Accountability in Indiana

Damian W. Betebenner
National Center for the Improvement of Educational Assessment (NCIEA)
Dover, New Hampshire
DBetebenner@nciea.org

September 25, 2014

Executive Summary

Indiana, like all other states, is engaged in numerous, substantive changes to its education system that are likely to have significant impacts in the coming years. These changes, often precipitated by state and federal policy include: The development and implementation of an A-F school accountability system, the adoption of state specific college and career ready standards, the development and implementation of an educator evaluation system, and the transition of the ISTEP+ state standardized assessment to one based on College and Career Ready (CCR) standards.

This report examines a thread that runs through many of these changes: the analysis of student growth based upon Indiana's state assessment (current the ISTEP+ and transition to the Career and College Ready Assessment). Indiana has been at the national forefront of calculating and using student growth. Beginning in 2008, the state began calculating student growth percentiles (SGPs) for diagnostic purposes and using them to identify areas of strength and weakness. Later, following the implementation of the state's A-F accountability system, the state included student growth as a prominent component of its school accountability system. Student growth is currently used in Indiana in the following areas:

1. School/A-F Accountability
2. Educator Effectiveness/Evaluation
3. School Improvement Plans
4. Charter School Accountability

As Indiana makes design modifications to its future accountability system, an important aspect of that design process is the recognition of the many different forms that accountability assumes in the state.

This report reviews the proposed changes to student growth in Indiana's A-F accountability system in both a state and national context. Based upon this review the following recommendations are made:

- Changes to the state's accountability growth calculations should be vetted in terms fairness and transparency.
- Changes to the state's accountability growth calculations should be vetted in terms of coherence with other state accountability uses of growth.
- Changes to the state's accountability growth calculations should be vetted in terms of consistency with state and federal policy.
- Changes to the state's accountability growth calculations should be vetted in terms of whether they can be calculated during and following the transition from ISTEP+ to CCR.

The Accountability Systems Review Panel was tasked with revising the states A-F accountability model including Indiana's growth model. As part of that year long review the panel analyzed numerous approaches ranging from strictly categorical to blended categorical/growth approaches. The strictly categorical approaches demonstrated unacceptably high correlations with prior achievement (and likely bias). The blended approaches showed better results that were consistent with results found in other states. Based upon these analyses, the blended option D demonstrated the best performance characteristics and best follows the above 4 recommendations.

As Indiana transitions to its new assessments and standards, the transition to a new accountability system should be completed in as seamless a fashion as possible so as to minimize disruption. To that end, the state will implement, for the last time, a targeted growth approach in the 2014-2015 school year followed by a transition to the new growth methodology (recommended here as option D) in 2015-2016. Currently, there are no insurmountable technical hurdles preventing the transition to occur. However, like in other states, the daunting issues related to assessment, standards, and accountability system transitions are about communication more so than technical. It is recommended that Indiana think broadly and long term about clear and effective communication during this multi-year transition to new assessments, standards, and accountability

A Summary of Growth Approaches

Over the last several years, the Indiana Department of Education has pursued am ambitious course of change in its testing and accountability system. From the A to F school accountability system, the development and implementation of a new educator evaluation program and new state content standards to the coming implementation of a new state assessment aligned to those standards in the 2014-2015 academic year, the scope of the changes in Indiana is immense. One common thread associated with all of these efforts is the calculation and use of student growth: Student growth is used in both school and educator accountability, is anchored to the performance and content standards on the current and future tests and is impacted by the switch from the ISTEP+ to the College and Career Ready Assessments (CCR) planed for the 2014-2015 academic year.

This report examines the analysis of student growth based upon Indiana's state assessment (current the ISTEP+ and transition to the Career and College Ready Assessment). Indiana has been at the national forefront of calculating and using student growth. Beginning in 2008, the state began calculating student growth percentiles (SGPs) for diagnostic purposes and using them to identify areas of strength and weakness. Later, following the implementation of the state's A-F accountability system, the state included student growth as a prominent component of its school accountability system. Currently, student growth is used for a number of different purposes in Indiana including:

1. School/A-F Accountability
2. Educator Effectiveness/Evaluation
3. School Improvement Plans
4. Charter School Accountability

Based upon recent and anticipated changes to the ISTEP+ student assessment and A-F accountability system, the state investigated various options regarding changes to its current analysis of student growth. Specifically, for the 2014-2015 academic year, Indiana will transition to a new assessment based upon a new set of state defined standards. Transitioning to a new state assessment based upon new standards presents several challenges for Indiana as it attempts to maintain an accountability system built upon an assessment and an associated set of standards that will no longer exist. In particular, states like Indiana, for whom student growth has become an integral part of their accountability system, are confronting several issues associated with maintaining student growth analyses.

As part of the current transition, the state has investigated a variety of growth methodologies that accommodate current policy initiatives, fit with previous commitments (e.g., ESEA

waivers), and accommodate the transition to new assessments based upon college and career readiness standards. Accommodating any one of these challenges is difficult, accommodating all of them is even more so.

This report addresses three broad considerations among different methodologies given the set of challenges the state is confronting:

Calculation: The ability to calculate the given quantities given the assessment transition coming in 2014-2015 Technical Quality: The technical quality of the growth quantities calculated assuming they can be calculated Communication: The communication challenges associated with the adoption of each method (e.g., consistency with currently used methods).

This chapter addresses three growth methodologies currently in use or under consideration by the accountability work group.

- 1 Year Projected (Currently used in A-F)
- Targeted Growth
- Categorical Status/Value Tables
- Student Growth Percentiles

Analysis of Growth Approaches

Data for analyses was supplied by the Indiana Department of Education and includes ISTEP results for ELA and Mathematics for Grades 3 to 8 across the past 2 years (2012 and 2013). Currently, the ISTEP assessment reports criterion-referenced achievement on a 3 level scale: Did Not Pass, Pass, and Pass +. As part of a re-analysis of student progress, the state is implementing a value-table model with 8 achievement levels and awarding points for transitions between achievement levels from the previous year to the current year. For an initial investigation, the state established cutpoints between the initial achievement levels by *equally* subdividing the current 3 achievement levels into 3, 3, and 2 sub-achievement levels, respectively.

The subdivision into 8 achievement levels was refined from an initial set of cuts due to too many students falling into just 4 of the 8 achievement levels. The 8 achievement level cuts decided upon reflect a better distribution of students across all of the 8 levels. The subdivision of the state's current 3 level scale into 8 subcategories was performed in an effort to create categories that contained substantial numbers of students as well as were not so close in terms of scale scores as to be subject to over interpretation due to measurement error. However, given that the ISTEP+ (and state tests in general) tend to have the majority of students scoring in the middle of the scale score/achievement level distribution, creating categories that were "wide" relative to the standard error of measurement was not always possible. The state should review the cutscores relative to the standard errors and make sure that the bands are not so narrow as to raise concerns about whether the categorization of students into the 8 categories is capricious. This issue might also arise with the Career and College Ready Assessment that the state will be implementing in 2015.

For the 2012 to 2013 growth analyses using the refined categories, achievement level transitions of 382,690 students were analyzed in mathematics and 380,818 in ELA in grades 3 to

4, 4 to 5, 5 to 6, 6 to 7, and 7 to 8. Tables 1.1 and 1.2 show frequency counts together with conditional probabilities (the probability of observing the 2013 achievement level given the the 2012 achievement level) for the data.

The row totals for both the mathematics and ELA result show much better dispersion of students across the 8 achievement levels than was previous reported. In terms of value table calculations, movement between the 8 categories will have at least some consequence given that are a meaningful number of students in all 8 categories. Students primarily remain in the same category or move to an adjacent category. It is much rarer to see student jumping more than 1 category from year to year. For example, in ELA of the 10,417 students starting in Did Not Pass 1, only 893 (8.6 percent) of those student moved to Did Not Pass 3 and even fewer moved to the Pass category in a single year.

2012 Achievement Level	2013 Achievement Level								Row Totals
	Did Not Pass 1	Did Not Pass 2	Did Not Pass 3	Pass 1	Pass 2	Pass 3	Pass + 1	Pass + 2	
Did Not Pass 1	5,274 (0.506)	3,710 (0.356)	893 (0.086)	434 (0.042)	85 (0.008)	15 (0.001)	3 (0.000)	3 (0.000)	10,417 (0.027)
Did Not Pass 2	4,371 (0.159)	11,266 (0.410)	6,263 (0.228)	4,582 (0.167)	820 (0.030)	135 (0.005)	8 (0.000)	3 (0.000)	27,448 (0.072)
Did Not Pass 3	1,376 (0.046)	7,692 (0.255)	7,956 (0.263)	9,972 (0.330)	2,606 (0.086)	548 (0.018)	55 (0.002)	13 (0.000)	30,218 (0.079)
Pass 1	1,002 (0.013)	7,962 (0.097)	13,323 (0.167)	33,151 (0.415)	17,673 (0.221)	5,854 (0.073)	818 (0.010)	130 (0.002)	79,913 (0.209)
Pass 2	142 (0.002)	1,546 (0.020)	4,203 (0.054)	21,860 (0.280)	26,903 (0.345)	17,956 (0.230)	4,481 (0.057)	979 (0.013)	78,070 (0.205)
Pass 3	33 (0.000)	297 (0.004)	917 (0.012)	7,925 (0.100)	20,347 (0.257)	29,556 (0.373)	14,233 (0.180)	5,849 (0.073)	79,157 (0.208)
Pass + 1	3 (0.000)	27 (0.001)	110 (0.003)	1,161 (0.026)	5,142 (0.117)	15,012 (0.342)	13,129 (0.299)	9,364 (0.213)	43,948 (0.115)
Pass + 2	0 (0.000)	4 (0.000)	8 (0.000)	151 (0.004)	1,069 (0.033)	5,724 (0.180)	9,595 (0.303)	15,096 (0.477)	31,647 (0.083)

Table 1.1: 2012 to 2013 cross tabulation frequencies and (conditional probabilities) of ELA achievement level progressions using the modified 8 achievement levels.

2012 Achievement Level	2013 Achievement Level								Row Totals
	Did Not Pass 1	Did Not Pass 2	Did Not Pass 3	Pass 1	Pass 2	Pass + 1	Pass + 2	Pass + 3	
Did Not Pass 1	4,038 (0.390)	3,976 (0.384)	1,361 (0.132)	782 (0.076)	144 (0.014)	35 (0.003)	12 (0.001)	1 (0.000)	10,349 (0.027)
Did Not Pass 2	3,367 (0.127)	9,072 (0.343)	6,246 (0.236)	6,114 (0.231)	1,274 (0.048)	321 (0.012)	47 (0.002)	4 (0.000)	26,445 (0.069)
Did Not Pass 3	1,044 (0.038)	5,397 (0.194)	6,495 (0.234)	10,239 (0.369)	3,529 (0.127)	882 (0.032)	149 (0.005)	15 (0.001)	27,750 (0.072)
Pass 1	620 (0.009)	4,942 (0.072)	9,606 (0.140)	27,894 (0.406)	17,524 (0.255)	6,702 (0.098)	1,316 (0.019)	80 (0.019)	68,684 (0.180)
Pass 2	110 (0.002)	840 (0.012)	2,682 (0.039)	16,749 (0.242)	24,776 (0.242)	17,781 (0.357)	5,726 (0.256)	685 (0.083)	69,349 (0.181)
Pass 3	19 (0.000)	152 (0.002)	517 (0.007)	5,580 (0.077)	17,300 (0.237)	27,284 (0.374)	17,746 (0.244)	4,265 (0.059)	72,863 (0.190)
Pass + 1	3 (0.000)	32 (0.001)	57 (0.001)	909 (0.014)	5,008 (0.080)	17,720 (0.281)	24,626 (0.391)	14,624 (0.232)	62,979 (0.164)
Pass + 2	0 (0.000)	3 (0.000)	4 (0.000)	43 (0.001)	438 (0.010)	3,438 (0.078)	12,332 (0.279)	28,013 (0.633)	44,271 (0.116)

Table 1.2: 2012 to 2013 cross tabulation frequencies and (conditional probabilities) of mathematics achievement level progressions using the modified 8 achievement levels.

In terms of moving up to the next category, students from an initial achievement level do not have a uniform chance of moving up. The SGP methodology (Betebenner, 2009; Betebenner, Iwaarden, Domingue, & Shang, 2014) includes a criterion-referenced component for calculating growth targets for students that indicate what level of growth (in the SGP metric) is required for students to reach pre-defined achievement outcomes in a specified amount of time. In the context of the current transition table approach to looking at growth, the implicit goal is for students to move up (at least) 1 achievement at a time. For example, a student currently in Did Not Pass 2 wouldn't be expected to move to Pass 1 in a single year but would be expected to move to Did Not Pass 3 in year 1 and then to Pass 1 in the following year.

To quantify the likelihood of this happening, SGP targets were calculated for all 2013 students with the goal of them reaching the next higher achievement level within the next year. Tables 1.3 and 1.4 provide descriptive summaries of the SGP targets for students in ELA and mathematics, respectively, based upon their initial achievement level. For example, in Table 1.3, the median SGP target for students with an initial achievement level of *Did Not Pass 3* was 56. This indicates that 50 percent of students starting in the *Did Not Pass 3* category had a 1 year growth target to reach *Pass 1* of less than or equal to 56 and 50 percent of students had growth targets of greater than 56. The 3rd quartile for the SGP targets for this group of students was 62.0 indicating that 25 percent of the students starting in the *Did Not Pass 3* category needed SGPs in excess of 62 to reach *Pass 1* — an high, but not unreasonable, rate of growth reached by approximately 2 in 5 students.[1]

2013 Achievement Level	2013 SGP Target Summary					
	Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
Did Not Pass 1	3	35	44	45.2	54	99
Did Not Pass 2	3	45	53	52.2	60	99
Did Not Pass 3	2	48	56	54	62	98
Pass 1	4	58	67	65.3	73	99
Pass 2	7	56	67	69.9	86	99
Pass 3	10	56	67	69.9	86	99
Pass + 1	18	64	76	76.2	89	99
Pass + 2						

Table 1.3: 2013 ELA SGP targets to next higher achievement level based upon initial starting point using the modified 8 achievement levels.

In general, tables 1.3 and 1.4 show, especially for students at Pass 1 or above, that reaching the next level is often an ambitious but not unreasonable expectation for the majority of students. This compare favorably to previous distributions where the transitions to higher achievement levels were often too steep. For example, in Mathematics (table 1.4) students starting at *Pass 1* have a median SGP target to reach *Pass 2* in 1 year of 57. That is, half of the students starting at *Pass 1* need to grow at or above an SGP of 57 to reach *Pass 2*.

Value Table Results and Comparison

Following recommendations made to the board of education in a presentation in November, 2013 and the establishment of value-tables from the working group in February, 2014, 4 families of value tables were investigates: A, B, C, and D. Value tables associated with families A and B represented were strict categorical approaches where points were awarded based upon

[1]Note that a percentile can be converted to a probability by simply subtracting the percentile from 100. For example, a growth percentile of 62 corresponds to a rate of progress reached or exceeded by 38 percent of students. That is, the probability of observing that rate of growth or higher is 0.38, approximately 2 in 5 would demonstrate that rate of growth based upon current rates of student growth observed in the state.

2013 Achievement Level	2013 SGP Target Summary					
	Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
Did Not Pass	2	28	36	35.7	43	87
Did Not Pass 2	4	34	44	43.7	52	96
Did Not Pass 3	3	32	42	43	54	95
Pass 1	7	46	57	56.7	67	99
Pass 2	5	49	59	59.4	69	99
Pass 3	5	52	63	63.4	75	99
Pass + 1	12	61	67	68.2	76	99
Pass + 2						

Table 1.4: 2013 Mathematics SGP targets to next higher achievement level based upon initial starting point using the modified 8 achievement levels.

transitions between the 8 achievement levels. Value tables C and D represented models that blended categorical transitions with within category growth. The models were introduced in order A, B, C, then D based upon recommendations of the panel to refine and modify analyses based upon empirical results. The 4 families of models comprises a comprehensive set of approaches for Indiana to consider in modifying its current growth model.

Table 1.5 shows correlations between status and demographic metrics commonly used to evaluate growth models and the value-table summaries derived from the four value-tables proposed to the working group. In general, the correlations from methods considering *only* transitions between categories (A3, B1, and B3) show moderate to high correlations with their results and the prior achievement of students. These correlations are above what one sees, in general, with other growth/value-added approaches and give an indication that the value-tables selected thus far (particularly value tables B1 and B2 which were favored by working group participants) are likely biased against schools serving lower achieving and higher poverty' students.

		2013								
		A3	B1	B3	C1	C2	D(A)	D(B)	D(C)	Median SGP
ELA	Prior Percent at/above Proficient	0.597	0.597	0.597	0.404	0.365	0.292	0.321	0.287	0.311
	Prior Mean Scale Score Standardized	0.545	0.545	0.545	0.382	0.336	0.265	0.294	0.277	0.281
	Percent Free/Reduced Lunch	-0.381	-0.381	-0.381	-0.283	-0.251	-0.198	-0.174	-0.175	-0.169
Mathematics	Prior Percent at/above Proficient	0.43	0.43	0.43	0.271	0.217	0.206	0.211	0.198	0.232
	Prior Mean Scale Score Standardized	0.463	0.463	0.463	0.321	0.266	0.223	0.241	0.253	0.265
	Percent Free/Reduced Lunch	-0.268	-0.268	-0.268	-0.189	-0.147	-0.133	-0.141	-0.114	-0.128

Table 1.5: Correlation between status and demographic summaries and value-table growth results.

The approaches that allow for within category growth to be demonstrated (C1, C2, D(A), D(B), and D(C)) perform better in terms of correlation with prior achievement and school poverty. The correlations demonstrated by the models C and D are in line with correlations

observed in other states and so likely possess little model bias. The reason for the decrease in correlation is because a significant portion of students (approximately 40%) do not change categories and that in turn leads value tables A and B to show higher correlations. It would be possible to re-work value tables A and B to give lower correlations, but to do so would require lower values being given to higher achieving students and higher values given to lower achieving students. This would likely lead to unacceptable value tables. Correlations such as these has generally been the reason that value tables are not widely used nationally.

If taken as fact, the results in Table 1.5 indicate that, depending upon the value table chosen, the Indiana schools showing the highest student growth are, to a greater or lesser extent, predominantly the higher achieving schools and the Indiana schools showing the lowest growth are the lower achieving schools. i In such a scenario it will be rare to find a high achieving school showing low student growth or a low achieving school showing high student growth. The validity of that finding needs to be empirically confirmed with Indiana stakeholders as to whether it reflects what they believe to be the case. Compared to results in others states, the correlations associated with value tables A3, B1, and B3 are higher than what one commonly sees and are likely the sign of model bias. The correlations associated with value tables C1 and C2 are high but not out of range of what is generally seen. As a point of comparison, the correlations between median SGP and the prior achievement and free and reduced lunch indicators are shown and are generally lower than those indicated by the value-table results which is consistent with what has been found in other states.

Summary

This chapter presents results associated with achievement level transitions using an equal achievement level partitioning of the current ISTEP 3 achievement levels to 8 achievement levels. The results suggest marked improvement over the first iteration of value-table results with a fairly uniform distribution across the 8 achievement levels that have been defined. In addition, reasonable yet ambitious growth targets for students to reach the next level were also shown for the newly defined 8 achievement levels. Depending upon the value-table chosen, high correlations between the school values derived from the value-tables and key indicators should be examined with regard to stakeholders understanding of schools' performance statewide.

Growth and Assessment Transitions in Indiana

Indiana's current approach to analyzing student growth for use in its A-F accountability system has been in used since 2010-2011 an involves establishing growth targets that are used in the subsequent year to determine percentages of student making their growth targets. This approach is often referred to as targeted growth. The approach is unique to Indiana and was the focus of the Accountability Systems Review Panel's growth analysis discussions. This approach was approved to be used in the coming year in Indiana's most recent waiver application.[1] In 2015-2016, it is expected that Indiana will transition to a yet to be approved (option D was recommended in the previous chapter) growth methodology to be used for A-F accountability.

Growth and Assessment Transitions: A Multi-State perspective

State assessment and accountability leaders share many concerns about the transition from existing state assessment systems to new assessments, prominently those produced by the Partnership for Assessment of Readiness for College and Careers (PARCC) and the Smarter Balanced (SBAC) consortia. While these leaders are looking forward to the forthcoming changes in curriculum and instruction, many worry about maintaining accountability systems across the assessment transition. There is no question that there are many issues and challenges that generalize across almost all states, but it is nonetheless critical to tailor advice to the specific contexts under which these accountability systems are operating. States transitioning to PARCC, SBAC, and other assessments will have unique transition challenges relative to the design of their current state assessment systems. One of the looming challenges is the use of student longitudinal growth information in their school and educator accountability systems, ranging from simple gain score models to more complex value-added and student growth percentile models. Recognizing the importance of context-specific factors in providing technical assistance about states accountability transitions, the Council of Chief State School Officers (CCSSO) convened separate groups of PARCC and SBAC states all using student growth percentiles (SGPs) as part of their current accountability systems. State leaders wanted to talk with technical advisers and officials from other states wrestling with the extent to which the transition would affect existing accountability frameworks. Since the underlying assessments used to compute growth percentiles will be changing in all of these states, this transition will also likely require states to determine whether previous assumptions and inferences about growth taking place at the classroom, school or district level can be supported or sustained. Six states from PARCC and 9 states from SBAC attended the two CCSSO meetings facili-

[1]The targeted growth approach used by Indiana during the assessment transition year will require an equipercentile concordance between the ISTEP+ and CCR tests to be constructed so that targets can be established on the new CCR tests. Besides the concordance, the methodology employed previously is identical to what will be used in 2014-2015.

tated by staff members of the National Center for the Improvement of Educational Assessment (NCIEA). The PARCC state assessment and accountability leaders who met in December, 2013 were from: Arizona, Colorado, Massachusetts, Mississippi, New Jersey and Rhode Island. The SBAC state leaders met in July, 2014 and were from: Hawaii, Idaho, Indiana, Maine, Nevada, Oregon, Washington, West Virginia, and Wyoming. In each of these states, student growth percentiles (SGPs) are incorporated into key leading or primary indicators in school and/or educator accountability systems.

Though there are similarities across the different states, the challenges of incorporating growth data across the transition will vary based upon policy requirements, different accountability system designs, field testing requirements, and different approaches used to evaluate growth at the various levels (e.g., classroom, school, or district) in each state. This paper highlights technical, practical, and policy considerations for using growth percentiles during the assessment transition period to support state accountability efforts, and recommends analyses and guidelines to help inform decisions on how best to use growth percentiles for accountability purposes in light of these considerations.

Although the recommendations in this paper are tailored to states that attended the CCSSO-sponsored growth transition meetings, many of these general approaches may be useful to other states whether they are transitioning to PARCC, SBAC, or something else, and whether they are using SGPs or other growth/value-added models. However, as we noted above, advice presented here should be considered relative to the unique contexts influencing each states system.

Uses of Growth Data for Accountability and Transition Conditions

Implementation Timelines

Based on input received from the 16 states who participated in both meetings, many states intend to move forward with using SGPs during transition. However, some states have already moved forward with decisions on not using SGPs for accountability. For example:

- Colorado intends to run analyses to determine whether the transition SGPs should be used for school and district accountability, but has already informed districts that they are not required to use the transition SGPs for educator evaluations.
- Rhode Island will calculates SGPs for diagnostic purposes but will not report SGPs for accountability since they plan to generate SGPs once they have two years of spring-to-spring test results.
- Washington intends to use SGPs for school accountability, but has decided to suspend the use of SGPs for educator or leader evaluations until the 2016-2017 school year.
- Idaho is considering suspending the use of SGPs for accountability for one year during the transition period.

For those states moving forward with generating SGPs for accountability purposes during the first operational year of both consortia (2014-2015), SGPs will need to be calculated using

their respective legacy state assessment results as prior scores in the SGP calculation. For those states able to calculate SGPs in 2014-2015, they may also choose to continue using prior scores from their state tests during the second year of operation (2015-2016) because of the added stability associated with including a second prior score in the model.

Figure 2.1 depicts the timeline associated with the transition from the ISTEP+ assessment to the College and Career Readiness Assessment. Based upon a thorough review of the timeline at the December Growth Transition meeting, one of the most consequential considerations of this timeline is when scale scores from the 2014-2015 assessment will be returned.[2]

[2]Other states from both the PARCC and SBAC testing consortia as well as states developing their own assessments like Indiana have very similar timelines as Indiana. Assessment timelines for these scenarios can be found in the Appendix beginning on i Page 28.

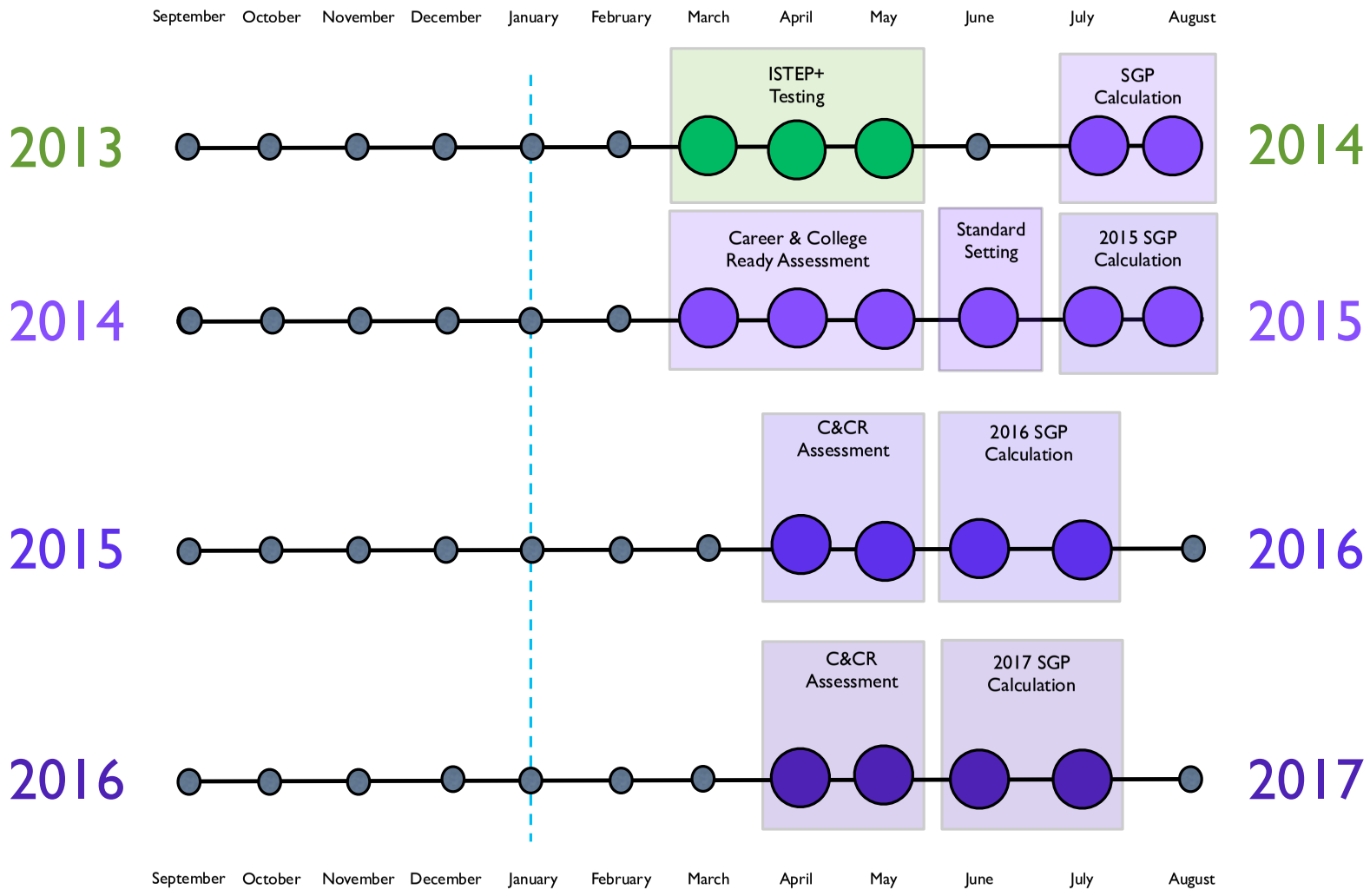

Figure 2.1: Assessment transition timeline for Indiana including growth calculation windows.

PARCC plans to release consortium-referenced SGPs in the 2015-2016 year and a few of these states may also consider the integration of those results in their accountability frameworks. At a minimum, the PARCC states will need a communication strategy to deal with two different sets of SGP results. All states leaders indicated that they are currently required to return assessment results and derivatives from those results such as growth (either by law or by rule), as early as August 1st. Though the timing associated with the 2014-2015 PARCC rollout is not final, current plans involve returning score reports as soon as possible after standard setting, which would mean sometime in early Fall 2015. However, in response to state requests, PARCC is considering various options for releasing scale scores, without achievement levels, prior to standard setting. At this time, the specifics of the available data and associated timeline await discussion within PARCC.

Figure 4.3 depicts the timeline associated with the transition to the SBAC assessment. A major challenge with the transition period identified during the meeting with SBAC states relates to the production of growth percentiles using 2013-14 results based on the field test designs deployed in several states.

Calculation of a student growth percentile requires, in general, scale scores from consecutive grades and years for a student. In contrast to PARCC states, some SBAC states have implemented different field testing strategies leading, in many cases, to missing data for students. SBAC field testing, like PARCC, returns no student-level data in 2014 with which to run growth analyses. This situation leads to three distinct outcomes occurring in SBAC states:

1. A state has a subset of students participating in the SBAC field test without double testing of the students on the traditional state assessment.

2. A state has all students participate in the SBAC field test without double testing of students on the traditional state assessment.

3. A state has a subset of students participating in the SBAC field test with double testing of those students on the state assessment.

Unlike status/attainment indicators, missing data impacts the calculation of annual growth of students in both current and future years. When the missing data are in the current year there is no current year data with which to calculate an SGP and in the next year the current years missing data become missing values for a prior year preventing the calculation of an annual SGP. Due to the considerable challenges of computing growth percentiles with these field test designs, we next weigh the implications associated with each of the three outcomes impacting various SBAC states.

Implications for Missing Data for Accountability in SBAC States

Outcome 1: A state has a subset of students participating in the SBAC field test without double testing of the students on the traditional state assessment.
This outcome, associated with the majority of SBAC states, leads to missing SGPs in both the current and coming year. In some states, with field testing percentages in excess of 25% of the students, this represents a large number of missing data points. In those states using growth for accountability purposes, these missing data can impact the accountability system in multiple ways including:

1. If the sample of students taking the SBAC field test is not random from the state, then the growth calculated for the remaining student participating in state testing will be systematically different making for different comparisons from previous years. Consideration should be given as to whether growth results should be re-normed or anchored to previous year's norms.

2. Going forward, two different cohorts of students will exist with different testing histories. At some point, the state will need to decide when to "sunset" the use of prior results from the state assessment and run analyses solely based upon SBAC results. At a minimum, this will occur when the state has two years of operational SBAC results available.

3. Districts, schools, instructors, etc. whose evaluation depends, at least in part, on student growth will not have that component available for the students with missing data. In cases of instructor evaluation, some instructors who previously received evaluations including the growth of students will not have that component available with missing data.

Outcome 2: A state has all students participate in the SBAC field test without double testing of students on the traditional state assessment.

This outcome where universal field testing occurs with no double testing, was indicated by only one attending state (Idaho), but may be a design in place for other SBAC states (e.g., CA). This outcome prevents any annual SGPs from being calculated but allows for bi-annual growth for all students with no bifurcation of the state population based upon them taking different testing tracks (see Option 1). The substitution of bi-annual growth for annual growth in accountability systems should be carefully considered. Impact studies using currently existing data can be done to help understand the impact of such a substitution. For example, states considering using SGPs that span more than a single year in lieu of annual SGPs can utilize currently existing data to calculate SGPs for that span (ignoring the annual structure of their current data) and investigate the relationship between these multi-year SGPs and their annual counterparts.

Outcome 3: A state has a subset of students participating in the SBAC field test with double testing of those students on the state assessment.

This outcome associated with a small minority of SBAC states but is most consistent with PARCC states, allows for annual growth to be calculated using the state assessment as the prior scores and the SBAC assessment as the current score in the transition year. This option is the easiest option in terms of continuing an accountability system including growth uninterrupted in the transition year as all the SGPs will be calculable barring any issues associated with the testing itself (e.g., floor and ceiling effects).

In general, the transition to SBAC and the existence of missing data in some circumstances will lead to missing SGPs which can impact accountability systems that rely upon that data. Understanding how wide spread such missing data is, whether such missing data is missing at random or systematically missing, whether bi-annual SGPs will be used in place of annual SGPs, how the missing data will impact the calculation of SGPs going forward, and how the data will be used based upon these considerations are some of the critical issues SBAC states need to consider.

	Growth Approaches Used in Accountability			Policy or Contractual Areas where Growth Results will be Impacted by the Transition				Growth Results Used in Accountability Systems Impacted by Transition				
PARCC States	Normative	Criterion-referenced	Baseline-referenced	ESEA Waivers	State Legislation or State Board	School Improvement Planning	Charter Contracts	Educator Evaluations	Leader Evaluations	School Accountability	District Accountability	School Improvement
Arizona	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓
Colorado	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓
Massachusetts	✓		✓	✓		✓	✓	✓	✓	✓	✓	✓
Mississippi	✓	✓		✓	✓			✓	✓	✓	✓	
New Jersey	✓			✓				✓	✓	✓	✓	
Rhode Island	✓			✓			✓	✓	✓	✓	✓	
SBAC States												
Hawaii	✓			✓		✓	✓	✓	✓	✓	✓	
Idaho	✓			✓	✓	✓	✓	✓	✓	✓	✓	
Indiana	✓	✓		✓	✓			✓	✓	✓	✓	✓
Maine	✓			✓						✓		✓
Nevada	✓	✓		✓		✓		✓		✓		
Oregon	✓	✓		✓		✓				✓	✓	
Washington	✓		✓	✓		✓				✓		
West Virginia	✓	✓				✓						
Wyoming	✓	✓		✓	✓	✓				✓		

Table 2.1: Uses and types of Student Growth Percentiles by state.

Summary of Growth Percentile Use for Accountability for Attending States

Prior to the meeting with PARCC and SBAC states, we collected information from each state to better understand the policy and implementation context in each locale. Table 2.1 summarizes information collected from each state from the two consortia in three areas: 1) current growth approaches used to support accountability decisions; 2) the policy and contractual areas likely to be impacted by changes in how growth will be used to inform accountability decisions; and, 3), the specific accountability systems in which growth is being used.

As indicated in Table 2.1, there are three primary approaches used by states to evaluate and report growth using the SGP framework in their accountability systems: norm-referenced growth (student growth percentiles), criterion-referenced growth (percentile growth trajectories) and baseline-referenced growth (baseline/anchored student growth percentiles) (Betebenner, 2009). We describe next how each of these approaches is typically used by states in the context of accountability.

Norm-referenced Growth refers to the common approach used by many states to summarize growth achieved at the school, classroom or district levels using the median or mean of the individual student growth percentiles calculated for each grade-by-content area norm group in the state. This approach was first developed in Colorado and subsequently extended to include criterion-referenced growth. Accountability determinations are then made by sum-

marizing the individual level SGPs using the median or mean and classifying the "average" growth achieved at each unit of analysis into various rating categories (e.g., low, typical, or high "average" growth). The number of classification categories set in accountability frameworks and the cut-points used to differentiate levels of growth performance vary across states.

Criterion-referenced growth is used to characterize the norm-referenced growth information provided by student growth percentiles relative to a criterion-based target defined most often by a proficiency standard of interest.[3] At the individual level, the result is a set of student growth percentile targets (target SGPs) indicating the amount of annual growth necessary for a student to reach or maintain a specified achievement outcome in a specified amount of time (e.g., proficiency within 3 years). This growth-to-standard implementation is often used by states to determine whether students are making growth sufficient to catch up/keep up to proficient status or move up/stay up to advanced status. Some states are using these data to calibrate their systems toward the goal of system-wide growth capable of getting students to career and college readiness as well as monitor movements made over time in the percentage of students falling into the catch up and keep up categories for school improvement planning and school accountability purposes.

Within the context of school and district accountability in several states, the median growth percentile (MGP) target needed for students at a school or district to achieve proficiency within three years is evaluated relative to the norm-referenced MGP achieved by each school or district. If the MGP achieved for a given school or district is below the MGP target needed to achieve proficiency, the school or districts growth is assessed using a higher bar or expectations for growth relative to other places where growth achieved either meets or exceeds the MGP target. A few SBAC states (e.g., Oregon and Idaho) are using this approach based on the accountability model originally conceived in Colorado, where one of two rubrics is used to rate a school or districts growth depending upon whether the school or district met the MGP target. In the rubric used for schools and districts where the MGP is equal to or higher than the target, the performance expectations have lower growth cut-points set relative to the rubric used for schools and districts where the MGP falls below the target.

Baseline-referenced growth is an approach used to evaluate growth by a student, school, district, or a classroom relative to a fixed "baseline" or "anchor" set of years established by the state. Growth percentiles for the majority of states are re-normed on an annual basis. Baseline-referenced data are used to enable the comparison of growth results over time and to permit the state to detect whether statewide growth is increasing from year to year. States using this approach anchor the norm groups each year to a baseline ensuring that the SGPs produced each year are interpreted relative to the baseline year.

Based on the information summarized in Table 1, all of these states face a set of challenges for using growth percentiles to support accountability inferences and decisions during their transition to the new assessments, whether they are using norm, baseline, and/or criterion-referenced growth:

Any accountability design changes implemented as a result of the transition may seriously affect multiple stakeholders at the state, district, and school levels.

Any potential changes to the design of accountability systems using growth data has clear implications for policy at both the state and district levels and presents challenges for commu-

[3]This proficiency standard could be set at various standards such as: proficiency, advanced, or partially proficient.

nicating those changes to all relevant users. There was universal agreement at the meetings that the foremost challenge was in communicating the changes to stakeholders: All states will need to communicate how growth will be used during the transition period and when certain approaches can be reinstated (e.g., criterion referenced growth or baseline referenced growth) at a future date. Furthermore, the transition period may have legal implications since the use of growth in the accountability reporting system prior to the transition will need to be re-purposed or re-framed. For some states, this would require approval at the legislative and/or board levels. In some cases, this would also require re-negotiating performance objectives specified in charter school contracts.[4]

During the transition period, all 15 states will need to consider how to interpret the growth achieved by students using the norm-referenced approach. Additionally, states using the criterion-referenced growth to standard approaches or anchoring growth data to a baseline year will likely need to modify or delay the use of those approaches during the transition period.

All states will receive student-level growth data from their respective consortium from the 2014-2015 school year allowing each state to generate SGPs using the new data in combination with their old state assessment data as priors. Many states have already used prior and current scores from different tests to generate SGPs, as in situations such when standards and assessments have been modified. However, the interpretation of growth results relative to the difficulty and complexity of the new content being assessed by the Common Core State Standards (CCSS) will need to be evaluated and considered by states opting to use those results to support accountability inferences. For the three PARCC states using or considering criterion-referenced approaches, none of these states are planning to move forward with this approach during the transition period. However, for the five SBAC states who are interested in maintaining the use of criterion-referenced growth approaches in their accountability systems, these states will be faced with the additional challenge of having to determine whether and when their criterion-based targets are defensible in light of changes in test content and standards for proficiency will likely become more rigorous. Finally, the baseline-referenced norm group approach used in Massachusetts and Washington will need to be suspended until the baseline years can be reset using data from the same assessment system. This will likely require waiting at least until the 2016-2017 to report baseline-referenced SGPs.

States must review and understand impact data to inform how best to use growth percentiles for accountability during the transition period. States must dedicate time and resources to carefully examine impact data in order to build defensible accountability systems.

Evaluating impact data to help support decisions on how to best use and report growth during the transition period is critical and will require both resources and time for states to consider accountability reporting options. States will vary in terms of the accountability design options they may pursue depending on what the impact data conveys when reviewing growth results generated prior, during and after the transition period, and likely after at least three years of PARCC or SBAC data are available.

[4]For more information about the broader issues associated with assessment transition and implications to accountability see: Domaleski, C. & Hall, E. (2014) Assessment Transition and Implications for Accountability. A white paper developed for the CCSSO Accountability Systems and Reporting State Collaborative on Assessment and Student Standards. Washington, DC: CCSSO. Available at: `http://www.nciea.org/publications-2/`.

As states receive the first year of PARCC and SBAC data, all states will need to conduct analyses to ensure that growth can be calculated from their prior tests to the current PARCC or SBAC test. In particular, a specific threat to the calculation of SGPs is the existence of prominent floor effects on these assessments. If a state's data yields a sizable percentage (¿ 5%) of students in any grade/content area receiving the lowest obtainable scale score, calculation of uniformly distributed percentiles will likely be difficult, if not impossible. Depending upon the extent of floor effects, the issue might be state specific or may apply consortium wide. Moreover, as the initial year's transition results become priors in subsequent years, issues associated with SGP calculation due to floor effects could continue as long as such data is used in the calculations.

Additionally, if the correlation of state-to-consortium assessment results is lower than expected (e.g., 0.6), the intended inferences about student growth from the legacy assessment to a consortium assessment may be suspect. One could argue that such less than perfect correlations may help identify schools or districts that have successfully implemented the CCSS compared with those that have not. While this may be true, it requires more than SGP results to support such an inference. Further, to the extent that CCSS implementation is associated with district wealth, incorporating transition growth results into accountability systems may threaten the credibility of the fairness of the accountability determinations because stakeholders might recognize that the growth results are related to the resources available to districts and not necessarily to educator or school effectiveness. On the other hand, it can be argued that the growth results may validly illuminate important differences in the fidelity of CCSS implementation.

States will be required to run analyses with regard to these and other potential threats. Further, the presumably more rigorous standards on the SBAC or PARCC assessment will likely lead to unique circumstances that will make the assessment transition period particularly challenging when using growth results to support accountability purposes. Unique local contexts and policy choices also complicate the decision to move forward with using or suspending the use of growth percentiles in a given state. For example, in the case of Massachusetts, all districts during the 2014-2015 school year have the choice to continue using the state assessment or administer the PARCC assessments. This policy will likely complicate efforts to compare and make inferences about growth achieved between districts, because the state will have to calculate two different sets of SGPs, each with a median of 50. That is, if the growth results for districts opting to use the PARCC assessments during the transition period are higher on average than the growth results for districts using the state assessments, comparative policy statements made about which schools exhibit higher growth performance cannot be supported since the underlying measures differ between those schools. Although these challenges apply to the 2014-2015 year, these challenges will not dissipate in the 2015-2016 year. Assuming all students take the same assessment, they will have two different sets of possible priors. This assessment choice policy also brings up the question of whether the same cut-points set in the current school and district accountability frameworks should apply to districts using different assessments.

In the case of other states such as Nevada, West Virginia, and Oregon, these states would prefer to maintain the use of criterion-referenced growth during the transition period, in particular to continue monitoring the extent to which disadvantaged groups are making adequate growth toward reaching proficiency targets. Although there are a few approaches that may

allow for these states to continue using the criterion-referenced approach, the inferences made about student growth using any one of these options carry a different interpretation than the inferences supported using student growth based on the legacy assessments. Even for places such as Colorado, where the decision was made to remove criterion-referenced growth during the transition period, removing this criterion piece from growth has brought up some concerns about tracking progress made by sub-groups since evaluating their growth relative to the proficiency standard served as an important indicator for tracking equity for these groups.

As indicated throughout this section, impact data will be especially critical for states to evaluate in order to support accountability design choices during and after the transition period. The next section outlines recommended analyses for using growth to support accountability decisions, and for potentially redesigning accountability frameworks prior to the transition.

SGP/Growth Analysis During and After the Transition Period

A common concern expressed by the 15 state participants regarding the assessment transition is whether it is even possible to calculate growth using two different tests. The simple answer is that it depends on the type of growth one wants to calculate. The change in assessments impacts the calculation of growth in different ways depending upon the type of growth to be calculated and the type of scale present in the state. However, as indicated earlier, considering the different approaches being used by each state to compute growth (refer to Table 2.1), the change in assessments may warrant consideration of different design options to adjust current accountability systems.

Overall, the technical issues associated with the calculation of growth are not as formidable as they have been previously with some states using SGPs. This is due to the fact that many of the technical challenges encountered have been addressed through the inclusion of new functions built into the SGP analytic software (Betebenner et al., 2014). A much bigger challenge is communicating results during the transition and deciding whether or how to implement different options for redesigning the current accountability frameworks. A key design consideration in the development of the Student Growth Percentile Model was to build a durable metric that could be maintained across assessment transitions. The extent to which substantial or minor adjustments must be made to accountability frameworks during the transition period again depends on the type of growth analyses used to support each purpose. In the sub-sections that follow, we discuss accountability framework dependencies relative to the growth approach used by each state.

Student Growth Percentiles

Despite the absence of a developmental (i.e., vertical) scale in the transition year with which to measure growth magnitudes, there is nothing to prevent the calculation of the norm-referenced SGPs. Norm-referenced SGPs measure student progress relative to academic peers who have the same achievement history as other students in the state. This is unlikely to be a concern since all students in each of the 15 states, except perhaps in Massachusetts, will continue to take the consortium assessment instead of their current state assessment.

Similar assessment transitions have occurred in numerous states in recent years. In 2010, Arizona changed its vertically-scaled math assessment to a new vertically-scaled assessment that did not correspond to the old scale. After some preliminary analyses showed that neither

assessment had significant ceiling/floor effects that could potentially undermine the calculation of SGPs, the state proceeded with calculating SGPs on the new system. Colorado recently transitioned from using the Colorado English Language Assessment (CELA) to the new Accessing Comprehension and Communication in English State-to-State for English Language Learners (ACCESS) assessment from the World Class Instructional Design and Assessment consortium. In order to justify the continued generation of growth percentiles for ELL assessments, the state ran analyses to check for ceiling and floor effects and to identify the extent to which prior CELA results were correlated with ACCESS results. The state also enlisted content specialists to evaluate and compare the similarities and differences in constructs assessed by both tests to justify inferences about growth in language development. Colorado has also been calculating growth from its state assessment to the ACT assessment, which has been administered universally in the state in grade 11 for a number of years in an effort to gauge growth toward college readiness. In essence, if the analyses and the review of constructs assessed by the different tests support the use of a different test as a prior, accountability systems developed to evaluate growth using a strictly norm-referenced approach do not require much, if any, additional work to modify the existing accountability framework.

When new assessments are introduced, the norm-referenced properties of SGPs will be preserved, insofar as the distribution of SGPs will be uniform for all students statewide. However, it should not be assumed that growth estimates are interchangeable with the growth outcomes students or schools would have received if the legacy state assessment were to continue. While the statewide distributions will be similar, the specific growth estimates that students and schools receive may well differ. This should be expected in the same way that one would not assume performance level classifications will be unchanged when a new assessment is introduced. For example, it stands to reason that a school with relatively strong performance on the new assessments will receive a more favorable growth estimate in the year following the transition compared to prior years, just as they would in any year with strong performance.

An accountability design consideration that should be investigated during the transition period with norm-referenced SGPs is the reporting of pooled SGP data across years. Pooling the SGP data entails reporting an MGP based on individual SGP results aggregated across two or more years. Colorado currently uses this approach for their accountability frameworks. The strategy of pooling data has the added advantage of stabilizing the year-to-year growth results reported for accountability purposes. It may also help guard against potentially large performance fluctuations during the transition. However, one area to consider when pooling data is that this approach will mask any year-to-year performance changes since the average performance achieved across years is reported.

Percentile Growth Trajectories

Percentile growth trajectories (also referred to as student growth projections) are a criterion-referenced growth metric on the same percentile scale as the SGP. They allow stakeholders to interpret the norm-referenced SGP in the criterion-referenced context of state assessments and determine whether the students growth is sufficient to put them on track to reach/maintain a desired level of achievement. Unlike student growth percentiles, which are retrospective and quantify what has occurred, percentile growth trajectories are forward looking and quantify what needs to occur in order to reach/maintain a desirable achievement outcome.

Percentile growth trajectories utilize the coefficient matrices calculated as part of the student growth percentile analyses to project forward along all 99 potential percentile trajectories for a student to determine what level of growth is required to reach various specified achievement outcomes (usually cuts between achievement levels) in specified amounts of time (annually going forward). As such, projections forward along a new test scale require coefficient matrices calculated relative to that scale. In the 2014-2015 transition year, dependent variable scores will be on the new (e.g., PARCC or SBAC) scale and independent variables will be on each states previous assessment scale.

As indicated earlier, several states in attendance (mainly SBAC states) are interested in maintaining the use of student growth projections or criterion-based growth as part of their accountability designs. These states will likely need to reconsider how growth can be used for accountability during the transition. As several states (in both PARCC and Smarter Balanced) rely upon percentile growth trajectories/student growth projections/student growth targets as part of their current accountability systems, one option for these states to consider is to statistically adjust the new test scale to create the best estimate of the prior test for purposes of continuing to produce comparable growth-to-standard outcomes. The downside of this approach is that the state may desire to model growth to a new and likely more rigorous standard and this approach simply delays incorporating the new expectation. On the other hand, it minimizes disruption in the model, which may be desirable if even for a limited time. For example, a state may be planning to implement substantial changes to the accountability model after two years of implementing PARCC or SBAC and wishes to defer major changes until such time.

Conversely, one could adjust the old test scale to create the best estimate of the new test to produce new growth-to-standard outcomes. These would not be comparable to the legacy growth-to-standard estimates, but would facilitate continued inclusion of this component in the accountability model. States that wish to incorporate the new standards in the model as soon as possible will likely find this approach appealing.

A number of approaches for producing statistically comparable scores include the use of an equipercentile concordance approach between the states assessment scale and the new PARCC or SBAC scale. Additionally, a number of equating approaches might be feasible if it is possible to have items in common between the tests or have a representative group of students take both tests. Using these equating approaches may provide a temporary solution for ensuring that the coefficient matrices used to project forward are based on an equivalent scale. This temporary solution may be applied in 2015-16, when more than one year will be available for a state on the new assessment scale, before the projections could revert back to using assessments from the same system. There are many technical reasons why we would not recommend this approach as a permanent solution to employ, but these are beyond the scope of this paper.

Another class of approaches involves identifying a suitable substitute for the growth target that does not require comparable test scales. To be clear, this alternative would break longitudinal comparability between the old method and the new method, but it may be an appealing option insofar as it allows the state to maintain criterion-referenced growth in the accountability system without statistically adjusting the old or new test scale.

Perhaps the most straightforward way to accomplish this is to identify a new growth standard, such as a school MGP target. This target could be identified through a combination of policy and data analyses to provide evidence that the target is sufficient for the purpose for

which it is designed. For example, the policy objective may be to select a target based on the legacy assessment such that the majority of students who are below proficient who grow at this rate become proficient in one year or some other time period. Given this definition, it would be straightforward to produce data that indicates the percent of non-proficient students in the prior year classified as proficient in the current year for each MGP.

Baseline Student Growth Percentiles

Massachusetts (as well as Georgia and Washington) makes extensive use of baseline-referenced student growth percentiles in their accountability systems. Student growth percentiles are a norm-referenced growth metric that are almost always normed using the most recent years data. As such, in a given year the median SGP for each content area/grade combination is 50. Some argue that re-norming growth every year in an accountability system results in a zero sum game. To rectify this situation, the Massachusetts Department of Elementary and Secondary Education established baseline growth norms against which future cohorts of students growth would be calculated. Based on the strong assumption of equivalence of grade and content area scales from year-to-year, the growth results can be anchored or baseline referenced such that the norm for the state is no longer fixed at 50 and that growth is interpreted relative to the specific baseline year of interest. Using this methodology, Massachusetts has been able to demonstrate higher rates of growth over time that are consistent with increasing the efficacy of the education system.

Considering the strong scaling assumptions employed to justify the use of baseline-referenced SGPs, we would not recommend using this approach in accountability frameworks until at least four years (two years to establish the baseline and two years to calculate growth against this baseline) of data are available from PARCC or SBAC to investigate how well these assumptions hold. In the specific case of Massachusetts, the baseline-referenced approach would need to be suspended for accountability purposes until more data from PARCC can be accumulated.

Recommendations

To summarize key recommendations as well as to highlight important analyses associated with the calculation of SGPs going forward, we recommend that states investigate performance during the transition year in the following ways:

1. For SBAC states with field designs related to Outcomes 1 and 2 (see pg.), the extent of the missing data will vary by state based upon their field testing plan. It is recommended that in states where missing SGPs will occur and those missing values will impact accountability systems, an inventory of the impact be conducted across the next several years to understand what will be available and how the state will accommodate the missing data across those years.

2. Examine the correlations between the pre- and post-scores used for students (e.g., a grade 5 math legacy assessment and a grade 6 PARCC or SBAC math assessment) and flag any correlations below 0.6 for inspection. Student growth percentiles can be calculated when scores are uncorrelated, but in such a situation, the conditional distribution associated with a fixed score is equivalent to the unconditional distribution, so the prior score

supplies no information. In other words, when scores are uncorrelated across years, SGPs would provide no information beyond status scores. When correlations fall below 0.6, it may indicate that there has been construct shift that should be recognized as part of SGP analyses.

3. For states that do not currently pool SGP data, explore the use of pooling across two or more recent years when reporting the norm-referenced results for accountability. As indicated earlier, this approach may mitigate the impact of performance fluctuations occurring during the transition period and may improve the precision of average growth estimates being reported for accountability purposes. This recommendation might be explored in general, even beyond the transition, to improve the stability of the outcomes.

4. Explore the use of results associated with replacing the growth target or developing comparable scales to evaluate whether this approach may support the continued use of student growth projections or criterion-based growth approaches in the accountability system during the assessment transition. As noted earlier, this should be considered a temporary patch until at least one additional year of data is available to support the projections using scores from just the PARCC or SBAC assessments. Examining the results to determine whether certain assumptions hold is critical, and although the specific analyses for evaluating these results are not addressed in this paper, we would strongly encourage states to ensure that they conduct these analyses first to determine whether the criterion-based approach can be maintained without interruption during the transition.

   Additionally, since it would be difficult to make the claim that these different assessments are written to the same content specifications, the equated scores reported need to be understood as establishing concordance between test scores rather than equating scores across assessments. Considering that growth inferences based on reported concordant scores do not share the same interpretation as growth inferences made using the legacy assessments, the issue of whether these equated transition scores should be reported for use in accountability systems will need to be discussed with stakeholders.

5. High school math course tests will be employed with the new PARCC assessments and this will require those states to investigate and find common course patterns across grades in order to generate SGPs. Since high school students can take different math course tests at different grades, states will need to evaluate the goodness-of-fit of the data relative to each of the common course patterns identified and ensure that adequate sample sizes are available to justify the selection of a specific pathway. Of course, this tends to be less of a concern for mathematics compared with subject areas such as social studies and science, because math course-taking tends to follow a predictable sequence, except for opportunities like honors and related courses.

6. Lastly, states should investigate the distribution of scores on the new assessment scale by grade and content area paying particular attention to percentages of students scoring at the lowest obtainable scale score (LOSS) of the scale. The non-parametric, quantile regression-based methodology underlying the calculation of SGPs and percentile growth trajectories is robust to most issues associated with assessment transition and the changing of scales. In other words, the SGP model is essentially agnostic to most linear scale

transformations (Betebenner, 2009) so that the new scaling, with possibly different interval properties than the legacy scale, is unlikely to affect calculation. However, if a large percentage (greater than 5%) of a cohort scores at the LOSS of the test, it becomes difficult to calculate SGPs for students that conform to a uniform distribution as is theoretically expected with percentiles. The non-parametric b-spline methodology associated with the SGP methodology allows the model to fit the data and recognize the LOSS and the highest obtainable scale score (HOSS) of the scale. Although corrections for LOSS/HOSS issues have been implemented as optional adjustments within the SGP package, if a preponderance of observations occurs at the LOSS for the new assessments, it may still be difficult to distinguish between student achievement outcomes leading to student growth percentiles when the data do not follow a uniform distribution. Given the design requirements for SBAC and PARCC, we do not anticipate LOSS to be a problem, but we are flagging this as a potential problem for many other states that are developing their own CCSS assessments. If significant percentages of students score at these levels, SGPs should be examined for model bias to ensure that disproportionate percentages of SGPs are not being assigned to students based upon floor effects of the test.

Bibliography

References

Betebenner, D. W. (2009). Norm- and criterion-referenced student growth. *Educational Measurement: Issues and Practice*, *28*(4), 42–51.

Betebenner, D. W., Iwaarden, A. V., Domingue, B., & Shang, Y. (2014). SGP: An R package for the calculation and visualization of student growth percentiles & percentile growth trajectories. [Computer software manual]. Retrieved from `http://cran.r-project.org/web/packages/SGP/index.html` (R package version 1.3-0.0)

Appendix

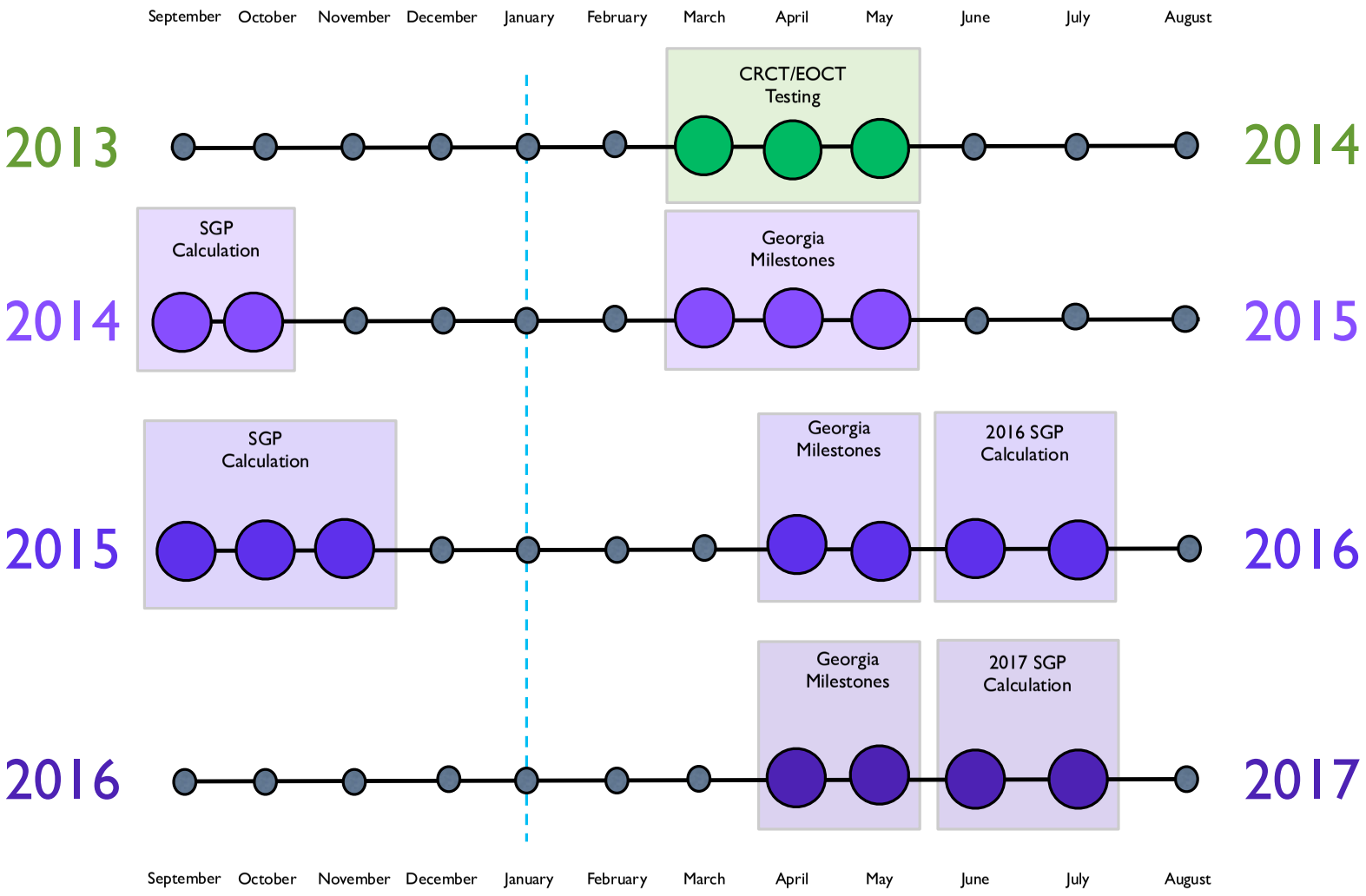

Figure 4.1: Assessment transition timeline for Georgia including growth calculation windows.

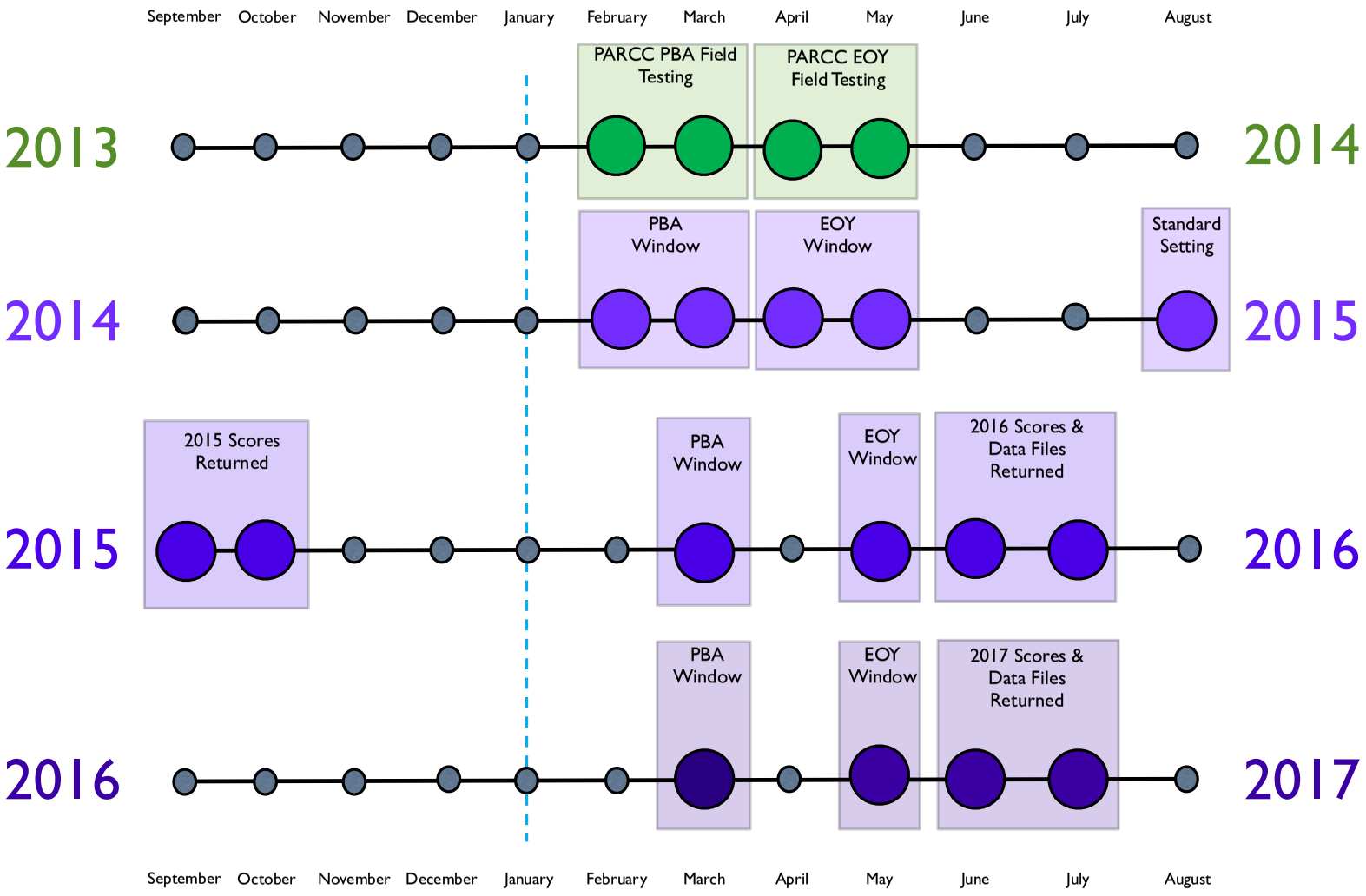

Figure 4.2: Assessment transition timeline for PARCC states including growth calculation windows.

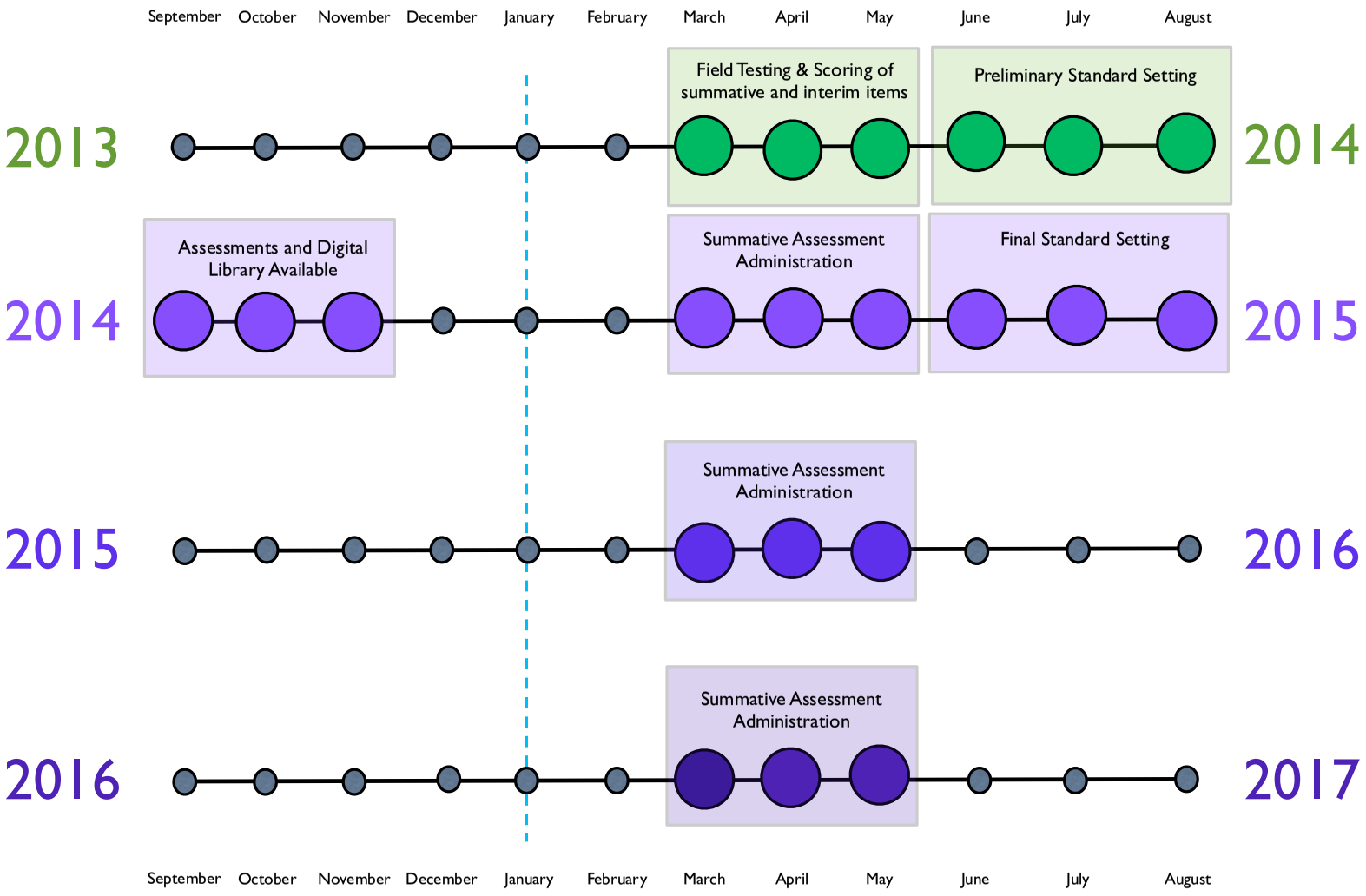

Figure 4.3: Assessment transition timeline for SBAC states including growth calculation windows.