

Calculating and Reporting Student Growth in Indiana

Methodological considerations for 2014-2015 and beyond

Damian W. Betebenner

National Center for the Improvement of Educational Assessment

Over the last several years, the Indiana Department of Education has pursued an 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 planned for the 2014-2015 academic year.

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:

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

This report addresses several growth methodologies under consideration in light of these challenges and provides recommendations on the method that best positions the state to make the best choice possible going forward.

Calculation

Following the June 16th, 2014 accountability workgroup meeting, a list of 5 methods for calculating student growth from the 2013-2014 ISTEP+ assessment to the 2014-2015 College and Career Ready assessments emerged as candidates for consideration:

1. 1 Year Projected (Currently used in A-F)
2. Targeted Growth
3. Categorical Status
4. Student Growth Percentiles
5. Improvement

Included in the list of 5 options are the currently used growth model (option 1) as well as the Categorical Status and Targeted Growth models currently under consideration to replace that model (options 2 and 3). Options 4 and 5 are listed as part of a larger inventory of approaches that are being used by other states that accommodate an assessment transition. This section discusses which of these options can be calculated based upon the upcoming assessment transition from the ISTEP+ to the 2014-2015 College and Career Ready Assessments.

Option 1: 1 Year Projected utilizes student growth percentile (SGP) calculations from the current year to establish scale score targets in the coming year allowing the state to determine the percentage of students in a school (or in any group of students) that met or exceeded their individual target. Because the 2014-2015 College and Career Ready Assessment will have a different scale than the current ISTEP+ assessment, the currently implemented analyses for 1 Year Projected growth cannot be performed. However, there are two ways that analyses almost identical to the current 1 Year Projected growth can be performed in the coming transition year:

- (i) Calculate scale score targets in the coming year using that data (instead of using the previous year's data) and then calculate the percentage of students in a school (or in any group) that met or exceeded their target.
- (ii) Calculate the scale score targets using the current data and use an equi-percentile concordance in the coming year to find the next year's scale score associated with the target so that a percentage of students exceeding their target can be calculated.

Option 2: Targeted growth utilizes student growth percentile (SGP) calculations from the current year to establish scale score targets in the coming year that indicate whether the student is on track to reach/maintain progress toward the next higher achievement level within three years. Because the 2014-2015 College and Career Reading Assessment achievement levels have not been established, these calculations are not possible until after two consecutive years of data from the new test are available to chart student progress toward higher levels of achievement.

Option 3: Categorical Status change, often referred to as value tables, is a method of examining student progress based upon changes in the performance levels of students from one year to the next. Using these observed changes, students are awarded points (more points for more desirable

transitions), and those points are used to create summaries for schools or other groups of interest. Categorical status change calculations are possible in the coming year and beyond with some caveats. Because the performance levels on the new test have not been established, values associated with the value-tables would not be able to be established until the new test is given and results are available for examination. That is, current work done on value-tables based upon ISTEP+ to ISTEP+ transitions, would not be applicable to the new assessment. Furthermore, in the 2015-2016 school year, the second year of the College and Career Ready Assessment, values associated with the value tables would have to be re-established again because students would no longer be transitioning from the ISTEP+. Because performance standards are expected to be higher on the new assessment, it is likely that the state would have to give points to students dropping achievement levels (e.g., from proficient in 2013-2014 to did not pass in 2014-2015) which might present communication challenges.

Option 4: Student growth percentiles are currently used in approximately 20 states that are transitioning to new assessments in the 2014-2015 academic year. These states include both PARCC, SBAC, and states developing their own assessments based upon career and college ready standards (e.g., Indiana and Georgia). Student growth percentile calculations are currently being performed by the Indiana Department of Education and are the foundation of the metrics discussed in Option 1 that the state is currently employing. SGP calculations are based upon an analytic technique that is invariant to monotone transformations of scale (Betebenner, 2009). As such, SGPs are scale neutral so that tests can be on different scales, and SGP calculations can still be performed. This is a desirable quality given Indiana's transition from the 2013-2014 ISTEP+ to the 2014-2015 College and Career Ready Assessment. States using SGPs often summarize growth using the median or the mean instead of the percentage of students above a cut as done in option 1. However, the percentage of students above a cut (like Indiana) is also popular and seen as more easy to understand.

Option 5: Improvement is not a student growth measure but instead looks at the change in the percentage of students deemed proficient in a group from 1 year to the next. After performance levels are set on the Career and College Ready assessment in 2014-2015, it will be possible for

the state to calculate improvement scores for schools and other groups of interest. Thus, calculation is possible, but the quantity calculated is not a student growth quantity.

All of the options presented are calculable by the state and require little to no outside assistance. Four of the five options are calculable in the coming transition year (1, 3, 4 and 5). Three of the options, 2 of which are calculable, require the state to first establish performance standards on its new assessment before calculation can occur (3 and 5). Of the 5 options, 3 of the options (1, 2 and 4) rely on the SGP methodology to calculate the quantity of interest. This fact highlights the reality that SGP analyses can be utilized in numerous ways by the state and provide great flexibility to both use and report multiple growth student growth indicators. More impactful differences between the 5 indicators emerge when looking at issues related to technical quality and communication which are discussed in the next sections.

Technical Quality

This section presents technical considerations of the 5 options listed that should be considered as part of Indiana due diligence regarding what methodological option provides the best fit. Considerations of technical quality are important but need to be balanced with communication considerations (discussed in the following section) in reaching a determination of what methodology is best for the state.

Option 1: The current 1 Year Projected growth calculations were implemented in a manner such that a target is established in the current year and students are examined relative to that target in the coming year. The reason for implementing this was to allow for the possibility of all students making their 1 year projected growth target and, by consequence, all schools or groups of interest having students making their targets. As such, results from a given year are not “zero-sum” and do not necessarily require anyone to be “low performers.” Potentially, everyone can hit their targets.

Of the two options considered to modify Option 1 for the current transition year, the equi-percentile concordance option allows Indiana to maintain the ability to have, in theory, all students above their targets in the coming year. In practice, this is unlikely to happen, which is

why most states just use the current year's norms in their analysis. However, communication considerations (to be discussed in the next section) are important in larger deliberations about which approach to take.

Option 2: Targeted growth, which is not calculable in the 2014-2015 academic year but would be in subsequent years, allows the state to identify individual growth targets for students that put them on track to reach or maintain state defined levels of achievement within established time frames. The growth-to-standard method avoids the short-comings of growth norms in that it unambiguously allows the state to tie student progress to desirable achievement outcomes. The use of targeted growth (which has been shown in the preliminary validation analyses for Indiana) often highlight the undesirable reality that growth targets for lower achieving students are much "harder" to accomplish than those for higher achieving students. As such, growth-to-standard results often show schools (or other groups) serving higher achieving students doing better than schools serving lower achieving students.

Note that this result isn't necessarily reflective of "bias" but instead highlights the reality of most state education systems: Students that start behind tend to stay behind and students that start ahead tend to stay ahead. Growth-to-standard methodologies formed the basis of many states' use of growth for school accountability under the Growth Model Pilot Program. States using targeted growth often supplement this indicator with a norm-referenced component (and SGP) so as to provide both a norm- and criterion-referenced growth indicator. Indiana is currently considering using value-tables together with targeted growth where targeted growth will highlight students making substantial changes within achievement levels.

Options 3: Value tables utilize the state defined achievement levels to examine student progress based upon transitions between these achievement levels from one year to the next. Based upon these student level transitions, the state awards points with more points being awarded for more desirable (i.e., upward) transitions. As part of Indiana's due diligence into growth methodologies, two sets of value-table analyses have been conducted to examine the outcomes associated with value-tables in the A-F accountability system as compared to historical A-F results.

In addition to the accountability based results, a number of analyses have been thus far conducted that looks at the relationship of student growth (i.e., value table results) relative to school characteristics like the percentage of free/reduced lunch students and the average level of achievement that the school serves. Such analyses serve to “smell test” the results so that the

		2013					
		A3	B1	B3	C1	C2	Median SGP
ELA	Prior Percent at/above Proficient	0.597	0.597	0.597	0.404	0.365	0.311
	Prior Mean Scale Score Standardized	0.545	0.545	0.545	0.382	0.336	0.281
	Percent Free/Reduced Lunch	-0.381	-0.381	-0.381	-0.283	-0.251	-0.169
Mathematics	Prior Percent at/above Proficient	0.43	0.43	0.43	0.271	0.217	0.232
	Prior Mean Scale Score Standardized	0.463	0.463	0.463	0.321	0.266	0.265
	Percent Free/Reduced Lunch	-0.268	-0.268	-0.268	-0.189	-0.147	-0.128

state understands what the particular model is communicating about schools in the state.

Table 1: Correlations between key school level characteristics, value-table results, and Median SGPs for 2013 ISTEP+ data

Table 1 shows 2013 correlations at the school level for key school level characteristics and value-table/SGP growth metrics in ELA and Mathematics. These correlations give a “system level” view of how growth relates to key school level characteristics across the state of Indiana. The table shows results for 5 different value-tables being considered (denoted A3, B1, B3, C1 and C3) as well as median SGPs. For example, across all schools in Indiana, the correlation between Prior Percent at/above Proficient (i.e., the entering level of student achievement as measured by percent at/above proficient) and A3 value-table growth is 0.597. This says that using value-table A3 results as a growth measure for school will show, in general, higher growth for schools serving higher achieving students and lower growth for schools serving lower achieving students. In other states investigating value-tables for adoption, high correlations have been of primary concern in setting values within the tables and whether to adopt this methodology.

As a basis of comparison, the correlation between achievement scores from one year to the next at the school level is approximately 0.7 to 0.8. For growth, a one year correlation of 0.597

between prior achievement and growth would be higher than what is observed in all states the author has worked with on growth. Other value-tables have lower values. Value table C2 have correlations of 0.365 and 0.217 for reading and mathematics, respectively, with prior achievement measured by percent at/above proficient and 0.336 and 0.266 when prior achievement is quantified using a mean standardized scale score. By comparison, the median SGP has correlations of 0.311 and 0.232 with prior achievement. It is important to emphasize that none of these results is necessarily “wrong” from a statistical standpoint. But from a practical standpoint, they may characterize the state’s schools in ways that are not commensurate with stakeholder beliefs.

The correlations between free/reduced lunch percentage at schools and the different growth metrics are also shown. Across all growth metrics, this correlation is negative, indicating that students attending schools serving higher percentages of free/reduced lunch students will tend to have lower growth. The growth metrics vary in terms of the extent of this correlation. The value-table metrics have higher correlation with percentage free/reduced lunch than does median SGP.

Option 4: Student growth percentiles and student growth projections are used by approximately 20 states for purposes ranging from teacher evaluation, school/district accountability, program evaluations, and diagnostic purposes. The analyses have been conducted in Indiana since 2008 and are currently used as part of the state’s A-F accountability system. As described thus far, SGP analyses have both a norm and criterion-referenced component allowing states to anchor growth to state defined achievement outcomes and also understanding what rates of progress are achievable by students in the state (i.e., ambitious yet reasonable).

Table 1 provides results showing correlations at the school level with key school level indicators. In general, the results show low to modest correlations between prior achievement and SGP based growth and low correlations between school poverty and SGP based growth. The adoption of any growth methodology should include due diligence where one looks at specific schools and makes sure the results are consistent with stakeholder understanding of quality.

Options 5: Technical considerations associated with improvement are difficult to describe as the quantity is not a growth quantity and so is qualitatively different and options 1, 2, 3, and 4. One likely reality associated with the use of an improvement metric is the reality that proficiency rates are likely to be much lower in the 2014-2015 academic year than they are currently. As such, it will be the “norm” for improvement rates to be very “negative.” The term “improvement” will be a misnomer in that in almost all schools, there will be a steep downward move in percent proficient. Establishing these norms may be difficult from a communication perspective. The author is not familiar with any state substituting improvement as a student growth metric in an accountability system and would advise Indiana to perform a number of analyses before using it as such.

Communication

Of the three considerations (calculation, technical quality, and communication), communication is likely the one with the biggest impact and the one that is most difficult to do right. A perfectly designed growth indicator combined with an ineffective communication effort will lead to failure; whereas, a less than perfect growth indicator combined with an effective communication effort will likely lead to (at least partial) success. Indiana has a number of years of experience in communicating results to stakeholders for both growth and status. The state should draw upon that experience to foresee and minimize the pain points that are likely to emerge due to the coming assessment/standards transition.

Consistency with previous efforts is an advantage if the goal is to not abandon the current approach. The benefit is that the state wouldn't have to “train” for another approach. As such, the “Year Projected” targets is the preferred method. If the goal is to “make a break” from the current approach, then options 3, 4, and 5 are available. Based upon work with the state over the last several years, the author has heard more dissatisfaction with how the growth metrics are being used than with the growth metrics themselves. That is, the way growth was weighted in the A-F calculations was of greater concern to stakeholders than the project growth metric itself. In stakeholder discussions about test-based accountability, the metric is often confounded with the use, and it is critical for the state to understand where the problem actually lies.

In terms of communication challenges, each option has its benefits/detriments given Indiana's past, current, and possible future use of growth. In reality, options 1, 2, and 4 are all based upon the same analytic technique (SGP). The scale invariance of the technique makes it flexible in how it can be “rolled

up” as the descriptive student growth percentile and/or “targets”. This flexibility provides options for providing multiple measures for stakeholders, some which could be used for accountability and others provided to further inform stakeholders. Student growth percentiles are currently used in approximately 20 states that are transitioning to a new assessment (and are the foundation of the metrics discussed in Option 1, 2 and 4). As noted, SGP calculations are based upon an analytic technique that is invariant to monotone transformations of scale (Betebenner, 2009). As such, SGPs are scale neutral so that tests can be on completely different scales, like the 2013-2014 ISTEP+ and the 2014-2015 College and Career Ready Assessment, and SGP analytics can still be performed without any impact upon the analyses. States using this metric often summarize growth using the median or the mean instead of the percentage of students above a cut as done in option 1. However, and as previously noted, the percentage of students above a cut is also popular and seen as more easy to understand.

Communicating results from SGP analyses benefits from the efforts of other states to use data thoughtfully as part of their assessment and accountability systems. For example, a number of states have developed video materials (that others states liberally borrow from) to explain student growth percentiles to non-technical audiences.

<https://www.youtube.com/watch?v=BDqj6t1UKYM>

<https://www.youtube.com/watch?v=dyArv7184ZY>

<https://www.youtube.com/watch?v=LWUJ2eKSwbM&list=UU9nZALC1qeHSdK8xzIXkVsw>

A consideration for Indiana is the extent to which their communication efforts can piggy back upon other states efforts. A goal behind the open source methodology of the SGP model is to allow open use of the resources associated with the analyses.

There is a common misconception with student growth analyses (and analyses in general), that effective communication is synonymous with explaining how a quantity is calculated. This misconception is similar to the misconception amongst math teachers that teaching a math concept means teaching how something is calculated. Understanding sometimes involves knowing exactly how something is calculated but often doesn't. For example, a middle school student doesn't necessarily know how the number pi is calculated but knows that it is a constant and that it is the ratio of the circumference to the diameter of a

circle. Similarly, a parent doesn't know how the height or weight percentile for their infant is calculated, but understands the quantity in that it tells them how extreme the height/weight of their child is.

One consequence of transition to a new assessment is that Indiana will no longer have a common scale that can be used to compare scores from one year to the next. In particular, the vertical scale that Indiana currently has with its ISTEP+ will not be carried forward so that scale score gains/losses cannot be computed. Without gains/losses, growth must be calculated using norm-based metrics that compare like students as they progress from the ISTEP+ to the Career and College Ready Assessment.¹

As Indiana confronts the communication challenges associated with the calculation of growth, it is critical to recognize that different stakeholder groups want different levels of detail. The majority of stakeholders are interested in a general understanding of the quantity (like the height and weight percentile for children); whereas, a minority want more detail on how the quantities are actually calculated. Multiple types of materials should be produced to address different stakeholder desires. Ultimately, communicating the results is about stakeholder understanding, which requires good “teaching” on the part of the state.

Summary & Recommendations

In considering which option is best given the current assessment transition, several considerations should be considered including: (i) ability to calculate, (ii) technical quality, and (iii) ability to communicate the results to the field:

(i) In terms of being able to calculate the relevant quantities, only option 2 (targeted growth) is unavailable. Option 1 Year projected quantities can be calculated in two ways. The presentation given June 16th indicates that the currently used “Year Projected” quantities can't be calculated. As mentioned above, they can't be calculated in exactly the same way done previously. However, very minor, and defensible, modifications to the current procedure can be used to calculate “Year Projected” targets.

¹ Even with a common (vertical) scale from year to year, growth norms are desirable in a similar way that height and weight norms are popular in communicating infant stature to parents.

(ii) In terms of technical quality, results associated with SGP analyses are more consistent with what is being seen nationally compared to those resulting from the value-table analyses. For example, some states with numerous large urban districts expect to see relationships between growth and poverty because of systemic problems that have plagued their states for decades. Indiana, like other states, is not immune to poverty, but whether it impacts educational outcomes in the same way will be indicated by the growth metric chosen. SGP analyses suggest that the impact of poverty on growth results in the state is much less than what is observed in other states using SGPs.

(iii) In terms of communication, consistency is likely an asset. If the goal is to maintain the current A-F approach, which benefits from the fact that one wouldn't have to "train" for another approach, then the "Year Projected" targets is the preferred method. If the goal is to "make a break" from the current approach, then options 3, 4, and 5 are available.

In terms of communication challenges, each option has its benefits/detriments given Indiana's past, current, and possible future use of growth. In reality, options 1, 2, and 4 are all based upon the same analytic technique (SGP). The scale invariance of the technique makes it flexible in how it can be "rolled up" as the descriptive student growth percentile and/or "targets". This flexibility provides options for providing multiple measures for stakeholders, some which could be used for accountability and others provided to further inform stakeholders.

Categorical Status change will require resetting of "values" two times (initially in 2014-2015 and then again in 2015-2016) which will require two value setting processes which may undermine credibility. In addition, in setting "values" it is likely that positive points will have to be given to students dropping in their level of achievement. This may present communication challenges as well that were also discussed at the June 16th meeting. If, as was being proposed, targeted growth also becomes a part of the categorical status model, then that would present another change to the model down the road.

Option 5, improvement, is not a growth model and the state would face the communication challenge of using a non-growth procedure in year 1 and, presumably, then switching to

something growth based for the following years. Improvement would also likely see considerable drops in percent proficient requiring the state to “endorse” such drops as OK in terms of accountability.

There is no perfect option for Indiana’s current situation. In laying out the the issues to be considered, in the opinion of the author, the most onerous is that of communicating the results to the field. The technical issues associated with the calculable options (1, 3, 4, and 5) are all easily surmountable. Because the state is already using SGP analyses as part of its A-F calculation and that those analyses can continue across the assessment transition and beyond, it seems desirable to continue those analyses but to, possibly, reconsider how they are weighted in the A-F accountability model. Student growth percentiles give the greatest flexibility for the state going forward in customizing indicators that have been validated nationwide for a variety of purposes including district, school and teacher accountability. This direction allows the state to pursue its currently approved A-F accountability system while considering refinements to the process as the new assessment comes online in the 2014-2015 academic year.

References

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