Recommendations for Defining and Measuring Growth:
Incorporating Work of the Growth Indicator Work Group

Summary

The Kentucky Growth Indicator Work Group, convened by Kentucky Department of Education Commissioner Wayne Lewis, was charged with recommending how growth should be defined and measured for the Kentucky school accountability system. The work group identified desired characteristics and principles of a growth model, and recommended specific ways to reflect those principles in practice. The Kentucky Department of Education and consultants operationalized those principles by creating specific tables and rules, and checked the results for technical quality.

Background and Process

During the process of gathering input during 2016-2018 for the new Kentucky school accountability system, "individual student growth" was consistently identified as a valued component. Guided by the input for several committees, the Kentucky Department of Education together with its technical consultants, developed a definition and approach for measuring growth that was accepted by the Kentucky Board of Education and incorporated into the determination of school accountability ratings and identification in fall 2018.

The key principles for the 2018 growth model included:

- Student growth was defined in relation to Proficiency: was a student "on track" to become Proficient within a certain number of years; "on track" to stay Proficient; or "on track" to become Distinguished
- "On track" was defined in terms of probable future performance, where probable future performance was defined as a projection two years into the future, based on the student's past performance combined with historical data from the state of likely performance of students with similar performance
- Schools should get credit for student progress, both moving up and moving down
- The credit should be rough proportional for each increment up or down, with the exception that students who start very high would not need to increment up to get credit
- The system should allow integration of growth results for students who participated with the Alternate and English Language Proficiency assessments
- The results should be of high technical quality, including useful, reliable, accurate, and resistant to gaming.
- The system and results should be credible, useful, and understandable by those whom the Kentucky school accountability system is intended to serve

The principles were embodied in a set of "value tables" that gave points based on where a student started and ended in terms of Kentucky's achievement levels. Novice and Apprentice were divided into two levels. The projected performance of students was based on the student percentile growth methodology and data from student performance on Kentucky's assessments from 2016 to 2017. (For more information regarding the growth model used in fall 2018, see value table below.)

Projected	Novice	Novice	Apprentice	Apprentice	Proficient	Distinguished
Current	Low	High	Low	High		
Distinguished	-1.50 (L)	-1.25 (L)	-1.00 (L)	-0.75 (L)	0.00 (K)	0.25 (K)
Proficient	-1.00 (L)	-0.75 (L)	-0.50 (L)	-0.25 (L)	0.25 (K)	0.50 (M)
Apprentice High	-0.75 (L)	-0.50 (L)	-0.25 (L)	0 (L)	0.25 (C)	0.75 (M)
Apprentice Low	-0.50 (L)	-0.25 (L)	0 (L)	0.25 (L)	0.50 (C)	1.00 (M)
Novice High	-0.25 (L)	0 (L)	0.25 (L)	0.50 (C)	0.75 (C)	1.25 (M)
Novice Low	0 (L)	0.25 (L)	0.50 (C)	0.75 (C)	1.00 (C)	1.50 (M)

The Kentucky Department of Education received many comments after the growth model and results were released in fall 2018. The main comments were:

- The growth model is difficult to understand and explain, especially "projected growth."
- The projected growth sometimes does not fit the observed performance of students; occasionally the mismatch is severe.
- Educators feel schools should be accountable for the results of students while in their schools, not performance that might happen in the future.
- The growth model is too tied to Proficiency, and somewhat duplicative of the Proficiency Indicator.

Commissioner Lewis authorized establishment of a Growth Indicator Work Group with the charge to consider options for defining and measuring growth, and to return a recommendation to him. The work group's recommendation should be completed in time to inform the Kentucky Board of Education's consideration of the regulation to guide school accountability ratings in fall 2019.

The Kentucky Department of Education recruited members of the Growth Indicator Work Group. Members included teachers, principals, superintendents, district assessment coordinators, and one data management specialist. (See Appendix A for list of members.) The work group met for approximately 6 hours each on October 24 and November 9, 2018 in Frankfort, Kentucky. The meetings were facilitated by Brian Gong, a consultant with the Center for Assessment who has national experience with growth models and detailed knowledge of the Kentucky school accountability system. Additional support was provided by Bill Auty, consultant who provides psychometric services to KDE, and Chris Domaleski, a colleague of Gong's who had also worked on the design of the Kentucky school accountability system.

The work group identified characteristics that a growth model should and should not have. The group also identified several principles of growth, and recommended specific ways to reflect those principles in practice. In making these recommendations, the group considered several examples that were generated by members of the work group, and also by the consultants. In its discussions, the group considered growth compared to a goal (e.g., proficiency), growth compared to where the student was previously (e.g., moving from Novice to Apprentice), and growth compared to norms based on how students performed in Kentucky. The group discussed

how these three approaches to measuring and evaluating growth interacted with different metrics (achievement levels, scale scores, and student growth percentiles). The discussions often returned to the desired characteristics, and how the growth model might be fair, actionable, understandable, reliable, non-gameable, etc.

Recommendations

Characteristics

The Growth Indicator Work Group identified several characteristics of what a growth model should and should not be. These characteristics reflected values that would inform the specifics of the recommended growth model. The characteristics identified by the group are shown in the table below.

Table 1: Desired Characteristics

Growth Indicator Desired Characteristics	
SHOULD BE	**SHOULD NOT BE**
Fair • Every school should have a chance to do well on growth • Growth should reflect what did happen, not what might happen • Equitable Actionable • Concrete goal known ahead of time for school improvement • Interpretable • Improved growth scores reflect "my kids and my school system" Different information than provided by other indicators in the assessment and accountability systems Predictive Simple Understandable, Explainable Usable/user friendly Reliable, consistent Accurate Non-gameable Leverage point Reasonable	Complicated Black box Projected Confusing Duplicated • EL students in accountability for both ELP and ELA

Principles

In their discussions, the Growth Indicator Work Group identified several principles that defined growth and how it should be treated in the accountability system. The work group separated the principles of what the growth model should be from discussion of how to implement it, because almost all of these principles could be embodied in multiple ways, i.e., using scale scores, achievement levels, etc.

Table 2: Recommended Principles

Growth Indicator Recommended Principles
1. Growth indicator should provide substantially different information than that provided by other indicators in the assessment and accountability systems 2. Growth should be based on observed (measured) student performance over time 3. No matter where student starts, comparable positive growth earns comparable credit 4. More positive change earns more positive credit 5. Learning more challenging content in a subsequent grade should be recognized as positive change, although the label may be the same (e.g., Apprentice in grade 3, Apprentice in grade 4; or 50th percentile in grade 3, 50th percentile in grade 4) 6. The growth system should focus on positive growth; all "negative growth" should be treated the same, regardless of whether a student declined more or less 7. Growth at extremely high and extremely low performances should be treated the same as other growth to the extent technically feasible 8. All students should be included in growth to the extent technically feasible, within the constraints of the law, including students with severe cognitive disabilities and English Learner students a. The growth credit should be as comparable across the different assessments as possible, e.g., maximum credit should be the same b. English Learner students should be included once, not twice, for growth in the accountability system 9. The system should be sensitive to growth, but the accountability results should be adequately reliable and resistant to being "gamed" 10. The system should be understandable, useful, and credible, especially with those who receive reports and who use the results to inform educational and/or policy actions a. The reporting scale should be clear (e.g., avoid negative numbers) 11. Growth should be included as an indicator in determining school accountability ratings; other aspects of growth performance might be useful to report as well 12. The growth model for Kentucky school accountability should be based on Kentucky achievement levels a. The Kentucky Department of Education should investigate the feasibility and desirability of using scale scores, particularly a vertical scale

The first 11 principles could be applied to almost any measurement metric—the group discussed scale scores, achievement levels, and a norm-reference growth percentile metric. Principle 12 identifies Kentucky's achievement levels as the measurement metric recommended by the work group.

All of these principles are allowable with federal law, with the possible exception of principle 7a, which was referred to KDE Legal Counsel for review. Note that principle 7a is about inclusion in accountability, not how growth is defined or measured. Although the Growth Indicator Work Group was not charged with discussing aspects of accountability, this was closely related to the principle of inclusion.

The work group discussed the advantages and disadvantages to growth of using achievement level categories, scale scores, and norm-referenced growth percentiles. Ultimately, the majority of the work group recommended using Kentucky's achievement levels (subdivided to improve sensitivity to change where feasible), as reflected in recommendation 11.

Using the principles and value tables discussed by the Growth Indicator Work Group, the Kentucky Department of Education and consultants created "business rules" to operationalize those principles, checked the results for technical quality and to help ensure the principles were met. The following table is based closely on tables that were discussed by the Growth Indicator Work Group, and reflects the principles. The table below differs from the tables given on charts at the Work Group meeting because the Work Group ended up agreeing on a set of principles that differed from any of the proposals made initially.

Table 3: Proposed Growth Value Table

Growth Indicator Value Table:
Points for student performance in Year 2, given performance in Year 1

		Year 2 Student Performance					
		NL	NH	AL	AH	P	D
Year 1 Student Performance	D	0	0	0	0	0	50
	P	0	0	0	0	50	100
	AH	0	0	0	50	100	150
	AL	0	0	50	100	150	200
	NH	0	50	100	150	200	250
	NL	0	100	150	200	250	300

NL=Novice Low; NH=Novice High; AL= Apprentice Low; AH=Apprentice High;
P=Proficient; D=Distinguished

How to read the value table: The table provides points assigned to combinations of performance in Year 1 and Year 2. The rows and columns are achievement levels (Novice, Apprentice, Proficient, Distinguished) where Novice and Apprentice have been divided into "Low" and "High." To find the assigned points for growth, find the student's performance level in Year 1 (left-hand column), and then read across the row to the cell under the student's performance in Year 2. The number in the cell is the number of points assigned for that student's growth.

How to use the value table to generate a school's growth score: For each student for whom the school is accountable for growth, use the table to find the number of points assigned for that student's growth. Sum the numbers and divide the total by the number of accountable students. Round to one decimal point for reporting.

Analysis

The Kentucky Department of Education and its consultants examined the proposed principles and specific means (i.e., the proposed and alternate value tables) for using the principles to generate schools' Growth Indicator scores. Key analyses focused on:

- Compliance of the Value Table with the Principles
- Interpretability of School Growth Scores
- Reliability and Sensitivity and
- Interaction with Other Accountability Elements

The analyses are discussed briefly below.

Compliance with Principles

The proposed value table complies with all 11 principles identified by the Growth Indicator Work Group.

1. Growth indicator should provide substantially different information than that provided by other indicators in the assessment and accountability systems – *The definition of growth as progress without relation to ending up proficient makes the growth indicator measure something quite different than what is reflected in the Proficiency Indicator. This is reflected in the correlation of about r=.3 between Proficiency and Growth at the school level using the recommended definition of growth, which is much lower than the correlation of r=.7 of the growth model used in 2018.*
2. Growth should be based on observed (measured) student performance over time – *The proposed value table is based only on observed scores. In particular, no projected scores (i.e., estimates of how a student is likely to perform in the future) are used.*
3. No matter where student starts, comparable positive growth earns comparable credit – *The number of points for moving up one or more levels is the same, regardless whether the students starts at Novice Low or Proficient.*
4. More positive change earns more positive credit – *Movement up for each additional level in the value table earns the same number of points, so moving up two levels earns twice as many points as do moving up one level, etc. This principle is kept, except where it conflicted with principle 8, where combining results from multiple assessments were considered.*
5. Learning more challenging content in a subsequent grade should be recognized as positive change, although the label may be the same (e.g., Apprentice in grade 3, Apprentice in grade 4; or 50th percentile in grade 3, 50th percentile in grade 4) – *"Maintaining" the same achievement level from one year to the next is awarded half as many points as moving up a level, and more points than declining.*
6. The growth system should focus on positive growth; all "negative growth" should be treated the same, regardless of whether a student declined more or less – *All declines in performance from one year to the next receive the same number of points—zero (0)—regardless of whether the student declined one performance level increment or more.*

7. Growth at extremely high and extremely low performances should be treated the same as other growth to the extent technically feasible – *Growth should be credited the same across the entire performance continuum (e.g., whether growth starts from the lowest or highest possible performance). However, there are instances where this is not technically feasible at this time. One instance is the case of extreme low: (Novice Low to Novice Low) is not given the same score points as is "maintaining" at other performance levels because it is not possible for students at Novice Low to decline a level, and a student could "maintain" at Novice Low by getting every assessment item incorrect. At the other end, maintaining Distinguished gets the same number of points as does maintaining at any other level. If it were possible to measure a meaningful decrease in Novice performance or a meaningful increase in Distinguished performance, then that could be reflected in the growth system. Currently it is not technically feasible to measure change at the very low or very high levels, as proposed for categorical change. Using scale scores might support more fine-grained determinations of growth, including at the very low and very high ends of performance, but there is not an available vertical scale that KDE or its consultants would recommend as technically feasible at this time. Note that a school with all its students at the Distinguished level will have a maximum Growth score of 50; the most meaningful interpretations will depend on using the Growth and Proficiency Indicators together.*
8. All students should be included in growth to the extent technically feasible, within the constraints of the law, including students with severe cognitive disabilities and English Learner students – *The proposed system allows inclusion of all students, including students with severe cognitive disabilities and English Learner students through a parallel growth table approach, customized to their specialized assessments. Whether such inclusion is supported by the technical quality of the assessments is discussed elsewhere. Students at the Distinguished performance level cannot show growth to a higher level; this is discussed more below.*
    a. The growth credit should be as comparable across the different assessments as possible, e.g., maximum credit should be the same. – *The growth indicator, if possible, should include growth for all students, regardless of which assessment they take. Kentucky has several different assessments: the general assessment, an assessment for student with severe cognitive disabilities, an assessment of English Language Proficiency for students who are English learners, and an assessment of English Language Proficiency for students with severe cognitive disabilities who are English learners. It is possible using the value table approach to give similar credit for growth as measured by the different assessments. It is the recommendation of KDE and its consultants that establishing a similar maximum growth score be one key way the growth scores are made comparable.*
    b. English Learner students should be included once, not twice, for growth in the accountability system – *English Learner students, after no more than two years of enrollment in a United States public school, are required by federal law to participate in two different assessments: an English Language Proficiency assessment and the state's academic assessments (e.g., Reading, Math). The proposed growth system allows results from both assessments to be included in a school's growth indicator score. Whether or not the accountability system should include both assessment results for English Learner students is an accountability policy matter and is discussed elsewhere.*

9. The system should be sensitive to growth, but the accountability results should be adequately reliable and resistant to being "gamed" – *The work group recognized that measures of growth must be sensitive to change, but not overly sensitive to the extent that results would reflect possible measurement error or have other poor technical quality. The group also discussed that the growth indicator should be resistant to "gaming." KDE and consultants reviewed the proposed growth measure to assure those using it that the growth measure would have desirable technical qualities, including reliability and resistance to gaming. In particular, members of the work group recommended that the Proficient achievement level be divided into "Low" and "High" sublevels, and similarly for the Distinguished level. The primary for this recommendation was to incentivize schools to work with all students, including higher performing students. KDE and its consultants examined the technical properties of the state assessments and recommend against dividing Proficient or Distinguished achievement levels at this time, since the measurement is less precise at those levels and the Proficient achievement level in particular is narrower, which makes it challenging to subdivide accurately.*
10. The system should be understandable, useful, and credible, especially with those who receive reports and who use the results to inform educational and/or policy actions
    a. The reporting scale should be clear (e.g., avoid negative numbers) – *The recommended growth value table and reporting scale are designed to be clear and useful; several changes were made based on lessons learned from the previous growth measure.*
11. Growth should be included as an indicator in determining school accountability ratings; other aspects of growth performance might be useful to report as well – *The Growth Indicator Work Group's primary responsibility was to recommend how to define and measure growth for school accountability, but the work group members who were present at its concluding meeting unanimously endorsed recommending including growth in the school accountability system for informing school ratings and other determinations. The work group recognized other growth information could also be valuably reported, even if not included in ratings.*
12. The growth model for Kentucky school accountability should be based on Kentucky achievement levels – *The proposed growth method, including value table, is based on Kentucky's achievement levels of Novice, Apprentice, Proficient, and Distinguished.*
    a. The Kentucky Department of Education should investigate the feasibility and desirability of using scale scores, particularly a vertical scale – *The work group recognized such a review probably could not be done thoroughly in the time frame required for application to the 2018-19 accountability determinations. KDE may wish to follow up on this in the future.*

Interpretation

There are many ways a school could get a growth score of 50. One way to get a sense of the type of performance represented by a score of 50 is that on average, a score of 50 is achieved by a school where, "All students maintained their achievement level from last year, on average, except if there were students who were Novice Low in Year 1; then of those about half on average moved up to Novice High." It is important that those who interpret the growth score

understand that a student who "maintained" a performance level while moving up a grade has learned a substantial amount, since the content standards are increasingly demanding.

Similarly, a score of 100 points for the school indicates that, on average, every student moved up one (half) level from the previous year, e.g., from Novice Low to Novice High, or from Apprentice High to Proficient. There are many ways a school could achieve a growth score of 100, by having some students move more than one level up and some move down, but on average the change in performance will be about one (half) achievement level.

Note that the growth score indicates how much and how many students grew, not where they ended up. Thus, one school could have a score of 50 with all students (maintaining) Apprentice Low, while another school could have a similar score of 50 with all students (maintaining) at the Proficient level. According to the principles laid out by the Growth Indicator Work Group, these two schools had the same relative growth, and so should get the same growth points. Of course, the two schools will differ quite a bit in their Proficiency scores.

There is not a straightforward, simple way to relate performance on the growth indicator with levels of proficiency since this growth model measures change without reference to where the student ends up, i.e., level of proficiency. In particular, there is no growth score that ensures adequate progress towards achieving the state's long-term goal of "reducing by 50% by 2030 the percentage of students who score below Proficient." A very good school growth score for a school with substantial percentages of students below Proficient, however, would be between 55 and 70. A school that maintained on average (except for some students at Novice Low), will have a growth score of 50; that school is not increasing the percentage of students scoring Proficient or higher. On the other hand, students who start below Proficient and have a growth score of 100 each year will be guaranteed to be at least Proficient within four years—so a school growth score of 100 is much higher growth than is expected by the state's long-term goals. A school score of 55 indicates that on average, 90% of the students maintained and 10% increased one level over the previous year.

Sensitivity, Reliability/Precision, and "Gameability"

The work group recommended using the most sensitive measure of growth possible that was also technically feasible and easily understood. Using scale scores—particularly changes in a vertical scale—has some advantages over using changes in achievement levels. For example, using scale scores would allow measuring growth anywhere along the performance continuum and not just in changes across achievement level boundaries; in particular, it might allow measuring growth at the very high and very low ends of performance. However, it was acknowledged that Kentucky's KPREP assessments do not include a vertical scale, and there is not currently an alternate vertical scale with demonstrated technical adequacy and ease of understanding. (The work group considered, for example, the Lexile and Quantile scales developed by MetaMetrics and reported along with KPREP scores.) KDE may wish to revisit developing and/or using a vertical scale to support measurement of growth in the future.

The work group strongly supported dividing the Novice and Apprentice achievement levels to be more sensitive to growth. The work group recommended also dividing Proficient and

Distinguished levels, and dividing Novice into three levels if technically feasible. KDE reviewed the precision of determinations and noted that dividing Proficient or Distinguished achievement levels would likely result in more misclassifications of students due to measurement error than is desirable. The same was true for dividing Novice and Apprentice achievement levels into more than two sub-levels.

Analyses done by KDE found that 1%-2.5% of students statewide had shown positive growth of three or more achievement levels (including the low/high subdivisions) in a single year. This is extraordinary growth. Students with such growth will raise the growth scores of their schools, particularly schools with fewer students. However, it is unlikely that such extraordinary growth will occur again the subsequent year. The implication is that a school's growth score may change somewhat from year to year. This is a function of the growth model being made sensitive to such unusual performances. Schools should not be surprised if there is variation in their growth scores from year to year; this directly reflects differences in observed performances from year to year. The work group considered some options to "dampen" the effects of such large observed improvements in performance (e.g., by awarding progressively fewer points for improvements of multiple achievement levels, or even capping the number of points), but in the end decided that principles 2, 3, and 4 expressed the group's values.

It was noted to the work group that the principle of giving differential credit to different positive growth, but giving the same credit to different negative growth makes this particular growth model more sensitive to positive growth and less sensitive to negative growth. As a result, schools may get similar scores although their negative growth is somewhat different because the value table does not discriminate among degrees of negative growth at the student level. An implication is that this growth model may not differentiate among schools at the lower end of growth, and therefore the schools' Proficiency and other indicators may carry more weight for making distinctions among schools with low growth. This is offset some by the lower correlation between this growth model's school scores and the schools' Proficiency scores.

While most members of the work group did not feel many Kentucky educators are likely to want to "game" the accountability system, the group members agreed that it was important to consider the possibility. It was pointed out that repeatedly moving a student down and then back up to the same level would earn the school positive growth points. However, moving the student down would reduce the school's proficiency score; manipulation of this type might result in a higher growth score but be offset by a lower growth score.

Interaction with Other Accountability Elements

This proposed growth model is weakly to modestly related to school proficiency scores in the same year. This indicates that the combination of growth and proficiency indicators provides more information about school performance than either one separately.

As mentioned above, Growth and Proficiency interact—both are needed in order to provide an accurate interpretation of school performance. In particular, because the growth indicator is focused on students maintaining or increasing their performance in terms of achievement levels, it is not possible to tell the proficiency level of a school by referring to the growth score—a

school with relatively low and a school with relatively high proficiency might have the same growth score, and two schools with similar proficiency scores might have very different growth scores.

Because of the principles chosen that underlie the growth model, the growth model is more sensitive to student changes in the positive direction and less sensitive to declines in student performance over time, i.e., “negative growth.” Specifically, any decline receives the same number of points, regardless of whether that decline is relatively smaller or larger. The implication of this is that when determining overall school ratings, growth will influence lower ratings proportionally less that it will higher ratings.

Appendix A: Members of the Growth Indicator Work Group

Mike Lafavers	Superintendent
Rob Fletcher	Superintendent
Jerry Green	Superintendent
Travis Hamby	Superintendent
Mike Borchers	Superintendent
Jana Beth Francis	District Assessment Coordinator
Teresa Nicholas	District Assessment Coordinator
Patrice Thompson	District Assessment Coordinator
David Meinschein	District Assessment Coordinator
Jeff Stamper	District Assessment Coordinator
Florence Chang/Joe Prather	Data Management Specialist
Twanjua Jones	Principal, Yates Elementary
Lorri Stivers	Principal, Shelby County West Middle
Nyree Clayton-Taylor	Teacher, Wheatley Elementary
Barry Baird	Teacher, Whitley County Middle