Title: Necessary Conditional Growth Percentiles – A way to connect conditional growth percentiles to achievement

Davie Store	William Sullivan	Joseph A. Marr	Eric G. Cronstrom
Central Michigan University	Michigan State University	Central Michigan University	Central Michigan University

Abstract

Measures for reporting student growth often suffer from two issues. Either they have no direct connection to achievement benchmarks and/or they are nothing more than achievement repackaged as growth. For accountability bodies, school level measures of adequate growth need several important features. First, they must describe growth that is 'adequate' for bringing students who have fallen behind up-to-standard. Second, they must be easy to understand and communicate. We derive a method from Thum (2011) for calculating Necessary Conditional Growth Percentiles (NCGP) and empirically test whether reaching a reasonable portion of necessary growth can, over time, predict benchmark achievement.

Objectives/Purpose

Many school accountability systems use both achievement and growth data to assess the quality of their schools. However, the amount of weight that is placed between achievement and growth differ among many accountability systems. In addition, while the connection between achievement and growth measures are clearly seen for many accountability systems, in others, there is lack of a clear-cut connection between achievement and growth.

Betebenner (2010) observed that the impact of NCLB upon research connecting large scale assessment outcomes and school quality has been profound. Current discussion often differentiates between accountability models/systems based upon status of the student and those that are based upon growth. Status models measure how students have performed at one point in time. With status models, students are put into categories also known as achievement levels depending on their level of mastery for the skills and knowledge being assessed. Such levels of mastery may include *advanced, proficient, partially proficient (basic),* and *below proficient*. Betebenner (2010) concludes that accountability systems constructed according to federal adequate yearly progress (AYP) requirements are based on a status model of achievement—annual “snapshots” of student’s achievement to make judgments about school quality. Specifically, NCLB requires schools to be evaluated based upon the percentage of students scoring at or above a “proficient” level of achievement. Unsurprisingly, such “snapshot” measures of student achievement have received heavy criticism from many stakeholders (Linn, 2003; Linn, Baker & Betebenner, 2002) for their failure to take into account the condition students were in before the current “snapshots” were taken. For example, an average-scoring student who started at the bottom may have achieved more than a high-scoring student who started out as a high performer i.e., a student moving from below proficient to proficient vs. another student who moves within a higher but same achievement level, e.g., from advanced level to advanced level. In such cases, the need to account for how far students have moved in terms of their performance in order to reach the status at which they are becomes contextually important. These inadequacies call for the use of growth models.

Student growth models therefore seek to qualify current achievement with respect to prior achievement. In practice, student academic growth has been defined as the change in achievement status between two or more time points. Growth models have been used to measure the changes that connect student scores over time. O’Malley, Murphy, McClarty, Murphy and McBride (2011) broaden the definition of growth models to include changes in proportions of students meeting proficiency from one year to the next, score changes made by students over two or more occasions and to measures that predict a student’s future scores. These authors elaborate that accurate or justifiable use of the term “student growth model” requires the satisfaction of the following conditions;

A) Scores can be mathematically compared from one occasion to another i.e., measured on the same scale

B) Scores can be connected to the same students over two or more occasions

C) Scores should show changes that indicate trait changes.

In addition to the inadequacies of status models stated above, another reason for the use of growth models is that they align well with one of the fundamental goals of education—student learning (O'Malley, Murphy, McClarty, Murphy & McBride, 2011). It is the goal of every educator to ensure that students are learning and that ultimately, students' level of skills and knowledge acquisition are improving over time. For example, when the NCLB legislation was used, states were required to show academic progress by bringing all students up to the proficient level on state tests by 2013-2014. This task required that accountability systems document how learning occurred on a yearly basis for all students (proficient or not) in order to determine whether schools are on track to achieving this goal. This could only be realized if growth models are used.

Further, unlike status models, growth models focus on the educational development of individual students regardless of whether they meet or do not meet proficiency targets. Parents, teachers and administrators are equally interested in the educational development of individual students and desire to understand whether observed student progress is reasonable and appropriate. This information is not reflected in status models that calculate the percentage of students meeting proficiency in a school.

There are many growth models that are used by different school accountability bodies. The key to using growth data appropriately and effectively is to properly align questions to be asked and the model that can derive such answers (Ernst & Wenning, 2009). Because growth analyses are designed to answer unique questions and accomplish different goals, growth models vary from one state to another or from one school accountability body to another. In general, in addition to using growth data to fulfil federal compliance requirements, states also use growth models for teacher evaluation, individual student comparisons, adequate yearly progress determinations and state accountability (Institute of Education Sciences-Statewide Longitudinal Data Systems Working Group, 2012). Currently, year-to-year gain models, level transition growth models, trajectory models, projection (prediction) models, student growth percentile models and value added models are the commonly used growth models among states that are fulfilling the purposes described above.

However, the ultimate goal for using growth models is to gauge how students are progressing towards the skills that are deemed necessary for different grade levels. While it is easy to connect growth to achievement using some growth models, the relationship between student growth and achievement is not clear-cut for other growth models. For example, it is not easy to make connections between normative growth and levels of achievement that students get to. As many have observed, the relative ranking of students based on percentage of students whom a student has outperformed on a test in itself does not provide an accurate measure of students' overall improvement. Ooms (2013) argues that students/schools are doing well only in a relative sense: they are doing well compared to other test-takers who took the same test—they may or may not be doing well in preparing students for college and career. It is likely that a student may have a higher percentile ranking when compared to a group of low achieving students and yet the same student may not rank higher when compared to another group of high achieving students.

This study uses Northwest Evaluation Association (NWEA) MAP Growth assessments and the growth projections and conditional growth percentiles that are derived using a multilevel growth modeling methodology (Thum and Kuhfeld, 2020), where a student's expected amount of growth depends on the starting score, grade level, subject and the weeks of instruction between two testing events. Many educators use NWEA growth data to assess how students are growing relative to the normative group. With the data, educators are able to make inferences about the relative ranking of their students in comparison to the normative group. NWEA periodically updates their norms to ensure that any changes in the performance of the population are reflected through the use of a new norming group.

As is the case with many growth percentile methodologies, connecting normative growth to achievement is important. Unfortunately, many users of NWEA growth percentiles have not taken the step to connect the growth that they observe from assessment reports to achievement. As a result, growth and achievement information are isolated – there is no real way of knowing how much growth will lead to reaching some specified levels of achievement e.g., proficiency on state standardized assessments. Stated differently, the inclusion of growth metrics devoid of any achievement context has resulted in questions about how much growth is required to meet a standard. This line of research is typically called "growth to standard" or "adequate growth." Conditional growth models tell us what we should expect of a student with relation to their peers. This has the advantage of setting reasonable expectations for student learning. For low performing students, this means comparing their growth to other low performing students. This often means that even if the student performs relatively well, they still do not catch up to the benchmark, just like their peer group. This is the deficiency in using conditional growth measures such as NWEA's Projected Growth to gauge school effectiveness. For most data users of conditional growth, their goal is to connect the normative growth to some standard such as state proficiency standards or college readiness benchmarks. To accomplish that, the use of adequate growth or necessary growth becomes vital.

In general, many accountability entities have developed protocols that attempt to incorporate achievement and growth in a multidimensional manner (status, trend, local comparisons). A priority in recent years is student growth. A leading reason for this priority on student growth is due to research that suggest that achievement rates (growth) varies largely between schools and not simply within schools (Bryk and Raudenbush, 1998; Attenberry and McEachin, 2020). A focus on growth means focusing on an outcome more insulated from contextual factors more likely to affect student achievement, like socioeconomic status.

Because reporting the amount of conditional growth a students or groups of students have year after year does not present any bearing in relation to the standards, even if the amount of growth is above typical, this paper presents a growth-to-standard measure which can be scaled to accommodate reasonable growth targets. This measure which we call Necessary Conditional Growth Percentiles (NCGP) performs better than Projected Growth at predicting students catching up to the benchmark over a three-year period.

Methods

NWEA's conditional growth percentiles are used as a basis for the development of our measure. These conditional growth percentiles are computed using the Conditional Growth Index (CGI), which is a standardized metric or a z score, that provides context for how much growth a student showed compared to his or her own projection. Precisely, fall to spring CGIs are calculated using the following three pieces of information: a) a student's fall to spring gain, 2) a student's growth projection based on the student's grade, starting RIT score, weeks of instruction and the standard deviation of fall to spring growth for a student's academic peers. Simply stated, CGI is a standardized difference between a student's observed gain and projected growth:

CGI = (Fall to Spring Observed Gain − Fall to Spring Projected Gain)/Standard deviation of growth

CGIs are then transformed into Conditional Growth Percentile (CGP) ranks to provide normative information about how student growth compares to that of other students across the nation. However, NWEA's conditional growth percentiles do not provide a sense of whether students' growth is adequate to meet grade level normative achievement benchmarks or other achievement standards, e.g., state standards on different state assessments. The current study therefore extends NWEA's conditional growth percentiles by examining Necessary Conditional Growth, which is the amount of growth needed for a student to meet grade level achievement norms. Similar to CGIs, Necessary Conditional Growth Index (NCGI) is computed as follows:

NCGI = (Difference between Fall score and grade level norm score − Fall to Spring Projected Gain)/Standard deviation of growth

NCGIs are also transformed into Necessary Growth Percentiles (NCGPs), similar to the same procedure as CGIs.

If a student's Conditional Growth Percentile is more than the student's Necessary Conditional Growth Percentile, then the amount of growth is deemed to be adequate. Conversely, if a student's Conditional Growth Percentile is less than the student's Necessary Conditional Growth Percentile, then the amount of growth is deemed to be inadequate.

However, fall to spring necessary growth (growth over one school year) may be too lofty a goal within such a short period, especially for very low achieving students who need to grow a lot in order to reach achievement grade level norms or standards. This is why many test users have started to look at necessary growth over a longer period (e.g., three years) to allow some room or a longer runway for students to achieve growth that will lead to meeting the required achievement grade level norms or standards at the end of the third year. Unfortunately, three-year norms for NWEA's MAP assessment do not exist at the time of this publication.

This method allows for a scaling similar to Thum (2011), where, a proportion (lambda value) of the difference between a student's fall score and the grade level norm score is deemed to be reasonable. Thum (2011) proposed .7 or .8 as reasonable amounts of proportion of growth. For our empirical investigation we adopt .7 as a lambda value. The formula for calculating reasonable NCGI is as follows:

Reasonable NCGI = Lambda (Difference between Fall score and grade level norm score) − (Fall to Spring Projected Gain)/Standard deviation of growth

Analysis

Projected Growth is the expected change in RIT score from one testing period to another. The expected score is derived empirically as the change in score for the median student out of a group of students who scored similarly in the past. Meeting Projected Growth can be operationalized as obtaining a change in RIT score greater than or equal to the projected growth value or as reaching a conditional growth percentile (CGP) >= 50. These measures are equivalent. NWEA sets the 'Projected Growth' score as the change in RIT that a student at CGP = 50 historically achieves.

Meeting Projected Growth does not mean, however, that growth is adequate. By adequate, we often mean at a pace that will enable a student to maintain an achievement benchmark or to reach the achievement benchmark in the future. The conditional growth percentile is not constructed with any direct reference to an achievement benchmark.

The first goal of this analysis is to empirically test whether meeting the Projected Growth target leads to students maintaining or reaching the achievement benchmark. For this analysis, we will define the achievement benchmark as a student attaining a CGP of 50 or greater. As stated above, this is equivalent to obtaining a norm referenced RIT score (within-subject, within-grade). That is, the typical (median) score for a student in grade *g* for subject *s*.

The second goal in this analysis will be to empirically test whether meeting an alternative measure of growth (NCGP) leads to students maintaining or reaching the achievement benchmark. For this analysis, we will use a lambda value of 0.7 for NCGP. That is, roughly equivalent to saying a student reached 70% of the growth necessary to reach the benchmark in the corresponding school year.

Further, we define Weak-Start and Strong-Start as students who begin the school year with a Fall MAP score that is at or above the previous grade level's normative benchmark. We do this because we typically think about students belonging to one of two camps, students who are learning at an appropriate pace for their grade level, and students who are behind.

Data Description

The analytic sample for this project is NWEA MAP results from 2016-2017, 2017-2018, and 2018-2019 school years for all schools in the portfolio of schools for a large mid-western charter school authorizer. The full sample contains 22,165 students in grades 3 through 8 for 41 schools. We want to analyze growth overtime. This requires that we restrict our sample to schools that are in the dataset for all three years. This results in 41 schools. We will restrict our analysis here to mathematics only. Restricting further for student's with observations in all three years leaves us with 3362 students in 33 schools. We further restrict for low counts within school. Four schools have less than ten students who have mathematics scores for all three years. Dropping these observations leaves us with 3344 students in 29 schools.

Cross-Section Analysis

For Tables 1, 2, and 3 (in Appendix), Projected Growth is the growth measure of interest. For Strong-Start students, meeting Projected Growth (CGP >= 50) virtually ensures that the student will meet the benchmark that year. For Weak-Start students, meeting Project Growth means that the student meets the benchmark about 23 to 30 percent of the time.

We conduct the same cross-section analysis for the NCGP measure (Tables 4, 5, and 6). Again, Strong-Start students are basically ensured to meet the benchmark when they achieve the NCGP growth measure. For Weak-Start students, approximately 50% of students who meet the NCGP also meet the benchmark. This is an improvement over Projected Growth of more than 20 percentage points.

Longitudinal Analysis

In this longitudinal analysis, Strong-Start is defined as being at or above the previous grade's benchmark in the first year. Met the growth measure (Projected Growth or NCGP) is defined as meeting the growth measure in all three years. Finally, meeting the benchmark is defined as meeting the benchmark in the final year. Thus, we can track students' progress over three school years towards maintaining or attaining the achievement benchmark.

For Projected Growth (Table 7), we note that for students who have a Strong-Start in the first year and who meet Projected Growth for all three years, nearly all meet the achievement benchmark in the final year. For Weak-Start students, meeting Projected Growth in all three years means that 54.5% of students meet the benchmark.

For NCGP (Table 8), a slightly lower percentage of students (97.4%) of Strong-Start students meet the achievement benchmark. However, 89.7% of Weak-Start students meet the achievement benchmark if they also meet NCGP in all three years. Note also that 'false positives' (students who do not meet the growth measure) yet meet the achievement benchmark are lower for NCGP vs. Projected Growth.

Sensitivity and Specificity

We can borrow the ideas of sensitivity (true positive rate) and specificity (true negative rate) that are commonly employed in fields such as epidemiology and biostatistics to assess our measures of growth. Sensitivity refers to the ability of a test to correctly identify individuals with a certain condition. Specificity refers to the ability of a test to correctly identify individuals who do not have a certain condition. These concepts are often used to evaluate the predictive quality of machine learning algorithms and statistical models.

We define a positive test outcome as a student reaching the NCGP growth target in all three school years. That is, necessary growth becomes our 'test' that a student is growing at an adequate pace to reach the achievement benchmark. For Projected Growth, we use the same logic. True positives will then be students who meet the benchmark in the third year. Table 9 contains the sensitivity and specificity calculations for each test. The first two rows apply to all students in the sample. The last four rows report results conditional on students' starting point.

Remarks and Conclusion

We note that NCGP as a test of students reaching the achievement benchmark performs much better than Projected Growth. In particular, the sensitivity (true positive rate) of NCGP for Weak-Start students (low performers) is a full 20 percentage points higher than Projected Growth. We believe that this 70% NCGP growth target is a reasonable target for schools to set for their lower performing students. These empirical results suggest that achieving this target overtime improves the chances that these students will catch up. School level reporting of the proportion of students meeting NCGP takes on valuable meaning.

In addition, the simplicity of aggregating this measure is appealing, especially to accountability bodies, such as, Charter School Authorizers. Performance reports that indicate the proportion of students who meet NCGP or who have met NCGP for consecutive years is simple to track and report. Improving this proportion becomes a reasonable target. More importantly, the need to connect norm based growth measures is increasingly in high demand now, especially due to the effects of the COVID-19 pandemic on learning. Student performance reports during the COVID-19 pandemic from Northwest Evaluation Association (Kuhfeld et al., 2020) and from Renaissance Learning (2020) show that there are significant learning losses due to the pandemic. As a result. Most accountability bodies will put more focus on growth in upcoming years, to ensure that students are quickly brought to pre-pandemic performance levels. It is therefore imperative to know how much growth it will take students to get back to required levels of achievement, e.g., state standards.

To demonstrate how the data can be used, figure 1 shows the distribution of students depending on where they started from and whether they met their projected growth using mathematics data from the 2018-2019 school year. When the amount of growth is set at 70% of the 'distance' between students' fall scores and the grade level normative scores it was observed that about 47% of the students met necessary growth. For all schools, the goal is to ensure that students who are already behind (low achieving students) are growing more in order to catch up. To that end, it is more concerning for educators when they observe that a school has a large percentage of weak start and low growth students (red section of the chart). Educators are encouraged to look for ways to push schools that have a lot of low achieving students to ensure that their students have more than projected growth (the blue section of the chart). Unfortunately, some students' achievement levels are way behind that even when they meet projected growth, they are not able to meet 70% of the distance to the target. Conversely, for schools with a lot of high achieving students, they are encouraged to make sure that their students are still making high growth to maintain their achievement levels – ensuring that they have few students in the yellow sections of the charts.

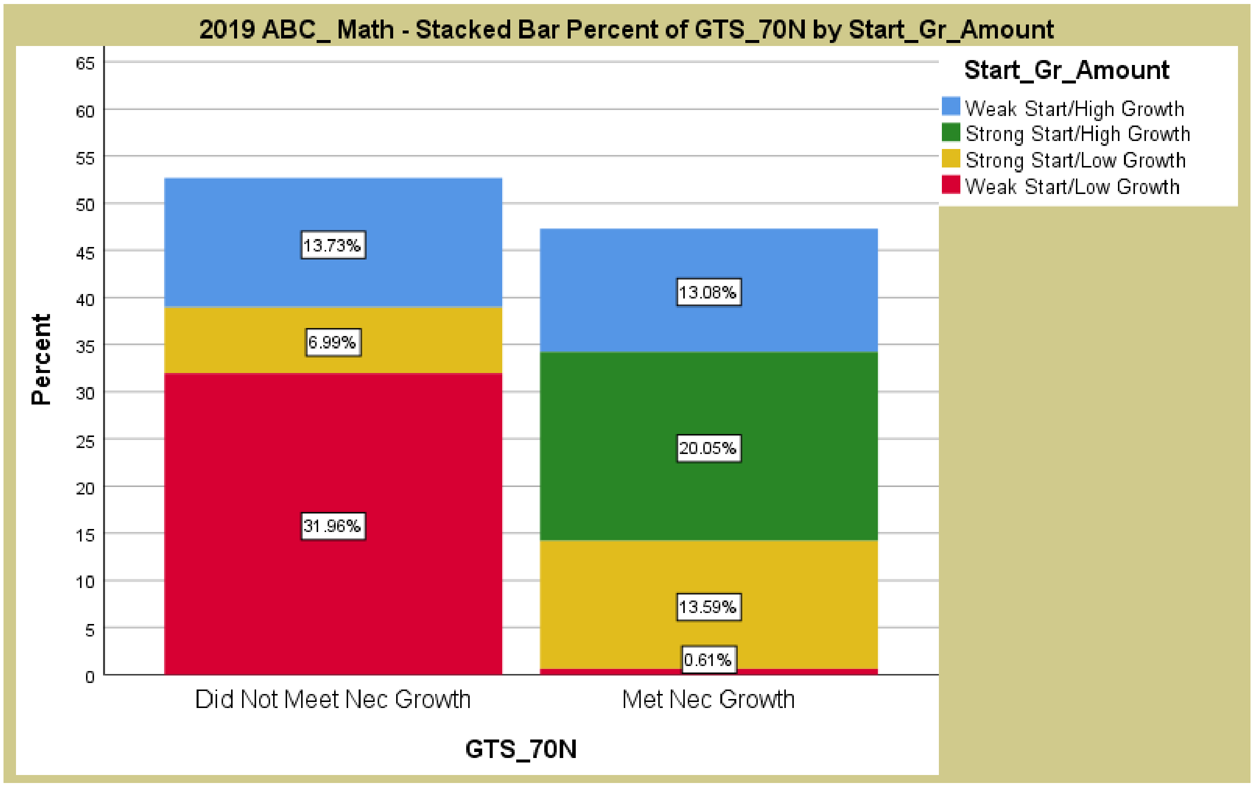

Figure 1. Percent of students in different categories when the proportion of growth is 70%

While an attempt has been made to connect growth to achievement (grade level performance on NWEA MAP tests), which is also normative in nature, there are two follow-up studies that could be done. First, because grade level achievement norms only show the average performance of a student in a specific grade level, there needs to be more research done to explore the specific skills that underline grade level normative performance. Second, the study could be expanded to use state assessment proficiency scores as a measure of achievement that can be connected to growth.

APPENDIX

Table 1

Cross- Section Analysis for Projected Growth Using 2016-17 Mathematics Data

Start	Projected Growth	Benchmark Not Met	Met	Sum
Weak	Not Met	100.0	0.0	100.0
	Met	70.1	29.9	100.0
Strong	Not Met	35.1	64.9	100.0
	Met	0.0	100.0	100.0

Table 2

Cross- Section Analysis for Projected Growth Using 2017-18 Mathematics Data

Start	Projected Growth	Benchmark Not Met	Met	Sum
Weak	Not Met	100.0	0.0	100.0
	Met	72.3	27.7	100.0
Strong	Not Met	36.5	63.5	100.0
	Met	0.0	100.0	100.0

Table 3

Cross- Section Analysis for Projected Growth Using 2018-19 Mathematics Data

Start	Projected Growth	Benchmark Not Met	Met	Sum
Weak	Not Met	100.0	0.0	100.0
	Met	76.8	23.2	100.0
Strong	Not Met	32.7	67.3	100.0
	Met	0.1	99.9	100.0

Table 4

Cross- Section Analysis for Necessary Growth Using 2016-17 Mathematics Data

Start	Necessary Growth	Benchmark Not Met	Met	Sum
Weak	Not Met	100.0	0.0	100.0
	Met	52.6	47.4	100.0
Strong	Not Met	89.4	10.6	100.0
	Met	6.2	93.8	100.0

Table 5

Cross- Section Analysis for Necessary Growth Using 2017-18 Mathematics Data

Start	**Necessary Growth**	**Benchmark Not Met**	**Met**	**Sum**
Weak	Not Met	100.0	0.0	100.0
	Met	46.1	53.9	100.0
Strong	Not Met	86.6	13.4	100.0
	Met	5.5	94.5	100.0

Table 6

Cross- Section Analysis for Necessary Growth Using 2018-19 Mathematics Data

Start	**Necessary Growth**	**Benchmark Not Met**	**Met**	**Sum**
Weak	Not Met	100.0	0.0	100.0
	Met	47.2	52.8	100.0
Strong	Not Met	68.2	31.8	100.0
	Met	2.3	97.7	100.0

Table 7

Longitudinal Analysis for Projected Growth

Start	Projected Growth	Benchmark Not Met	Met	Sum
Weak	Not Met	85.4	14.6	100.0
	Met	45.5	54.5	100.0
Strong	Not Met	22.8	77.2	100.0
	Met	0.5	99.5	100.0

Table 8

Longitudinal Analysis for Necessary Growth

Start	Necessary Growth	Benchmark Not Met	Met	Sum
Weak	Not Met	89.4	10.6	100.0
	Met	10.3	89.7	100.0
Strong	Not Met	49.2	50.8	100.0
	Met	2.6	97.4	100.0

Table 9

Sensitivity and Specificity Rates

Group	Test	Sensitivity	Specificity
All	Projected Growth	20.5	95.5
All	Necessary Growth	68.9	97.9
Strong-Start	Projected Growth	18.0	99.6
Strong Start	Necessary Growth	77.0	91.4
Weak-Start	Projected Growth	27.2	94.9
Weak-Start	Necessary Growth	47.6	98.8

References

Atteberry, A., & McEachin, A. (2020). Not Where You Start, but How Much You Grow: An Addendum to the Coleman Report. Educational Researcher. Advance online publication.

https:// doi.org/10.3102/0013189X20940304

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

Betebenner, D. W., & Linn, R. L. (2010). Growth in student achievement: Issues of measurement, longitudinal data analysis and accountability. Retrieved from Center for K-12 Assessment and Performance Management Website: http://k12center.org/publications/next_gen_exploratory_seminar.html

Bryk, A. S., & Raudenbush, S. W. (1988). Toward a more appropriate conceptualization of research on school effects: A three-level hierarchical linear model. American Journal of Education, 97(1), 65–108.

Ernst, J., & Wenning, R. (2009). Leave No Charter behind: An Authorizer Guide to the Use of Growth Data. Authorizing Matters. Issue Brief. *National Association of Charter School Authorizers*.

Institute of Education Sciences-Statewide Longitudinal Data Systems Working Group (2012). Issues and Advice from the States: A Guide of the Statewide Longitudinal Data Systems Grant Program. Retrieved from National Center for Education Statistics Website: http://nces.ed.gov/search/?output=xml_no_dtd&client=nces&site=nces&sitesearch=nces.ed.gov%2Fprograms%2Fslds&q=advice+from+the+states&submit=Go

Kuhfeld, M, Tarasawa, B., Johnson, A., Ruzek, I, & Lewis K. (2020). *Learning during COVID-19: Initial findings on students' reading and math achievement and growth.* NWEA.

Linn, R.L. (2003). Accountability: Responsibility and Reasonable Expectations. *Educational Researcher, 32(7), 3-13.*

Linn, R. L., Baker, E. L., Betebenner, & D. W., (2002). Accountability Systems: Implications of Requirements of the No Child Left Behind Act of 2001. *Educational Researcher, 31(6), 3-16.*

O'Malley, K.J., Murphy, S., McClarty, K.L., Murphy, D., & McBride, Y. (2011). Student growth models. Retrieved from Pearson Assessments Website: http://www.pearsonassessments.com/research/researchpub/researchlist.topic.growthModeling.html

Ooms, A. (2013). Voices: The problem with growth scores. Retrieved from Education News Colorado Website: http://www.ednewscolorado.org/voices/voices-the-problem-with-growth-scores

Renaissance Learning (2020). *How Kids are Performing: Tracking the impact of COVID-19 on Reading and Mathematics Achievement.* Special Report Series, Fall 2020 Edition (https://www.renaissance.com/how-kids-are-performing/)

Thum, Y. M. (2011). Measuring student growth and achievement against college readiness benchmarks and the ACT. National Charter Schools Institute.

Thum, Y. M., & Kuhfeld, M. (2020). NWEA 2020 MAP Growth achievement status and growth norms for students and schools. NWEA Research Report. Portland, OR: NWEA.