E3 ALLIANCE
EDUCATION EQUALS ECONOMICS

White Paper

Measuring and Improving Student Learning: The Case for 3D Growth

3D GROWTH

Authored by: *Simon Tidd, Ph.D.*
Mark Bond, M.S.

May, 2017

E3 ALLIANCE
EDUCATION EQUALS ECONOMICS

Executive Summary

The assessment of student performance is a core component of the educational system. Measuring performance at the student, teacher, and school levels provides vital information about how well individual students are learning and how to improve the systems guiding and staff teaching them. Performance standards based on passing rates on single point-in-time assessments (e.g., STAAR and EOC) are insufficient and mask important information about student learning (i.e., growth), focus disproportionate attention on students at the threshold of passing ("bubble kids"), and are strongly correlated with student characteristics such as income level, making comparisons between teachers and school tenuous at best.

The use of Student Growth Percentiles (SGP) offers a powerful alternative approach to conceptualizing and supporting student learning. Broadly similar to pediatric growth charts, SGP compares the performance of a student to a group of her or his academic peers – students with a similar history of achievement. Each student is assigned a percentile rank indicating how well she or he performed relative to this peer group. Students with higher SGP scores performed more strongly compared to their peers, while those with lower SGP scores less so. A key advantage of SGP scores is that growth can meaningfully be assessed regardless of absolute performance. As a consequence, poor performing students who are making great strides can be identified as can strong performers who may not be reaching their full potential. In addition, because SGP scores are only weakly correlated with student characteristics (e.g., income level) it is possible to compare outcomes of students by teacher, program, or school and consider differences in growth no matter the characteristics of a given classroom or school.

SGP scores by themselves, while powerful, do not provide the framework or tools for systems improvement and effective student intervention. To provide that support, E3 Alliance has integrated growth scores into a set of comprehensive professional development workshops and data visualization tools called **3D Growth**. The workshops bring together campus and district leadership to develop an understanding of academic growth scores within the context of existing performance measures (i.e., STAAR/EOC) while simultaneously learning powerful planning and process mapping techniques with which they develop data driven implementation plans for targeting high-impact areas of need to improve student performance. Participants also learn how to integrate student level growth scores into a broader Response to Intervention (RTI) framework.

As part of the **3D Growth** support provided by E3 Alliance, participating districts receive growth scores at the student, teacher, and campus levels as well as benchmarking comparisons to other campuses throughout their region.

Measuring and Improving Student Learning: The Case for 3D Growth

> *"High-quality assessments are essential to effectively educating students, measuring progress, and promoting equity. Done well and thoughtfully, they provide critical information for educators, families, the public, and students themselves and create the basis for improving outcomes for all learners."*
>
> *- U.S. Department of Education, Every Student Succeeds Act Fact Sheet, 2017*

The Traditional Approach to Measuring Student Performance

The assessment of student performance is a core component of the educational system. Measuring performance at the student, classroom, school, district, state, and national levels provides vital information about how well individual students are learning and how to improve the systems teaching them.

Traditionally, performance has been measured by testing students on discrete assessments.[1] In Texas, STAAR tests and End of Course (EOC) exams are familiar examples, used to measure student essential knowledge and skills at the end of most academic years. These 'high stakes' assessments have been granted a critical role in student advancement and graduation as well as the evaluation of campus and district performance across the State of Texas.

Performance standards based on passing rates on single point-in-time assessments suffer from a number of important limitations, especially when attempting to understand the impact of teachers and schools on learning.

When focusing on the performance of a particular teacher, school, or district, student achievement data is typically reported in terms of the percentage of students who meet one or more performance thresholds (Neal and Schazenbach, 2010). Teachers, schools, and districts with larger percentages of students who achieve a standard are viewed as performing better than those with smaller percentages. There are several

[1] Our focus here is on summative assessment rather than formative assessment, which is used primarily for real-time feedback to adjust instruction.

important reasons why this approach to measuring performance and setting accountability standards is problematic.

Drawbacks of a Threshold Only Approach

First, from a statistical standpoint, important information about the variation between students in a class is lost. In a school with relatively uniform student needs and testing outcomes near the "meets grade level" threshold, curriculum and support needs may also be quite uniform. In a school with diverse testing outcomes and scores across the testing spectrum, curriculum and support needs may also be far more diverse… but the number of students meeting the "meets grade level" threshold may be equivalent. From an accountability standpoint, these schools look the same: an equal percentage of students *met standard*. However, from a school improvement standpoint, the story is very different. Students in the first school are likely more uniform in their needs with respect to curriculum and supports while students in the second school may need substantial differentiation to support students across the full range of performance.

The Texas Education Agency does provide additional thresholds (*Masters*, *Approaches*, and *Did Not Meet Grade Level*)[2] that provide more nuanced information about the distribution of student performance within a school or district. However, for students at the extremes – *Masters* or *Did Not Meet Grade Level* – the reporting of student percentages limits our knowledge of how strongly or poorly these students are performing.

Second, the average passing rate of a school is very highly correlated with the overall income level of students attending the school. As a consequence, observed differences in school effectiveness are difficult to separate from the challenges facing varying student populations.

Third, from a practice standpoint, the use of performance thresholds in accountability may place undue focus on so-called "bubble kids," students who cluster just above or below an important cut point. While performance targets are fundamentally important in terms of setting baseline levels of knowledge and skills for student success, so is the learning of all students no matter where they fall in terms of 'passing the test.' The pressure to focus resources narrowly on a small set of students – sometimes with the goal of improving their performance even marginally - for the purpose of campus and

[2]*Note*. Beginning with the 2016-17 school year STAAR and EOC assessments, the Texas Education Agency switched to a new set of performance categorizations (Did Not Meet Grade Level; Approaches Grade Level; Meets Grade Level; Masters Grade Level). These categorizations are expected to remain in place until 2023-24 when the phase-in of new standards is complete. At this point, STAAR and EOC performance will again be based on two thresholds – *Meets* or *Masters Grade Level*.

district accountability is an unfortunate reality, albeit one that runs counter to the intent of the Federal *Every Student Succeeds Act*: "improving outcomes for all learners."

And finally, threshold-based achievement measures are at their core broad and generalized, representing the end product of learning. As such, they are poorly suited to understanding the real challenges individual students, schools, and districts face in teaching and learning. To use a specific example, two students start their 7th grade school year at two different schools. The first student struggled through 6th grade reading while the second student easily scored at the mastery level. After taking the 7th grade reading STAAR, both students received a score in the master range. From an accountability perspective, the outcomes look the same – both students performed very well. From a learning perspective, the story is more complicated. Both students mastered the 7th grade content, a wonderful outcome. However, the first student had to 'cover a lot more ground' to get to that point. At the teacher and school levels, the first student represented more of a challenge, and a much greater success story, given where she started than did the second student.

A Cautionary Note on the STAAR Progress Measure

In Texas, students also receive a second measure of learning, the STAAR Progress Measure. From year to year, student knowledge is expected to grow at a certain pace. This expectation is reflected in increases in the performance standards for each assessment. When a student's change in test score is equal to or outperforms the change in grade level expectations, she or he is described as having *expected* or *accelerated* growth. Students whose growth lag behind the increase in test difficulty are described as having *limited* growth.[3]

A challenge presented by this "progress measure" is that all growth may not be equal. At earlier grade levels, it may be easier to 'catch up' students to their peers because there is less conceptual material that may require scaffolding. At higher grade levels, it may be the reverse; prior poor performance might reflect more complex skill and knowledge deficits that slow a student's progress relative to her or his peers. Additionally, there is evidence that shows, from the reading literature in particular, that learning can be exponential, that is, growth is heavily dependent on where a child starts. The larger a students' early vocabulary, the easier it is for her or him to read. As reading comprehension improves, the student is able to acquire a larger vocabulary outside the classroom (Wagner, Torgensen, and Rashotte, 1994). Reading improves vocabulary which improves reading, and so on. This feedback loop can lead to exponential rather

[3] See *Calculating the STAAR Progress Measure* (Texas Education Agency, 2017) for a detailed explanation of how STAAR the progress measure is calculated for individual students.

than linear growth, or put simply, absolute growth is easier for students who have a leg up at the beginning.

The purpose here is not disentangle all the ways in which student growth may vary, rather it is to highlight that without clear expectations for what is typical versus atypical growth it is difficult to compare student to student, teacher to teacher, campus to campus or district to district. The STAAR Progress Measure, because it cannot assess what is typical or atypical growth across the full range of student performance, cannot provide an accurate measure of true student growth.

Contextualizing Student Growth

There are a wide variety of ways in which student academic growth has been measured. Simple difference or gain score approaches, such as the STAAR Progress Measure, describe growth in terms of absolute movement along a scale. As we noted above, growth of this type is difficult to interpret because 'strong' growth for a struggling student may look very different than 'strong' growth for an average or above average student.

More sophisticated statistical techniques have been developed that use information about a student's prior academic performance to provide a standard against which current year performance (i.e., growth) can be assessed. These approaches, broadly classified here as either *Value Added Models* (VAMs) or *Student Growth Percentiles* (SGP), differ in their goals.[4] VAMs were designed with the aim of *evaluating the 'value'* of individual teachers (or campuses) through their contribution to student growth. The methodology behind SGP was designed to provide an easily understandable *normative description of growth* experienced by an individual student. These design choices fundamentally impact the ease with which teachers, principals, and district administrators can use growth data to improve student outcomes.

Value Added Models

Value Added Models (VAMs) have increased in popularity over the past several decades as one method for measuring student growth. VAMs are used primarily to assess

[4] A number of detailed classification schemes exist for growth models (see Castellano and Ho, 2013, for example). Here we distinguish VAM and SGP broadly from measurement approaches that do not use predicted performance as a way of contextualizing present performance. Residual gain models use much of the same underlying statistical framework – growth expressed in terms of the difference between predicted and actual performance – as VAM models and are not considered separately here due to their limited use in an accountability context.

how well individual teachers and schools perform relative to peer educators and institutions.

At their core, VAMs are statistical models of growth that control for a variety of student level factors (e.g., prior test scores; income level) that correlate with future academic performance, as well as characteristics of teachers (e.g., tenure; certification type) and schools (e.g., size; climate) that also impact learning. Once a model has been specified, any remaining differences in student performance are attributed to either teacher or school effects. VAMs are superior to simpler measures of growth (e.g., STAAR Progress Measure) in that they assess performance relative to what was predicted rather than against an arbitrary standard. There are, however, important limitations associated with the use of VAMs for measuring student growth.

VAMs are based on the premise that all important factors related to student performance have been measured and are included in the model. In statistical terms, VAMs use the error variance – student performance relative to that predicted by the model – to measure teacher or school level effects. Realistically, it is impossible to anticipate let alone measure for all of the factors that may impact student learning, such as circumstances experienced in the student's home environment or the unique social dynamics of each classroom. As a consequence, VAMs are vulnerable to model specification errors that may conflate growth estimates with other unmeasured factors.[5]

VAMs are also based on the assumption that growth is linear; that future performance is a simple function of prior year performance. In reality, we often have no strong reason for assuming a linear relationship other than a desire for model parsimony and the constraints of the statistical technique. The degree to which growth is nonlinear will introduce bias into the growth estimates produced by this technique.

Perhaps the biggest limitation of VAM approaches to assessing student growth is how poorly suited they are to clearly describing growth at the student level. VAMs were designed, as the name conveys, to measure the relative *value-add* of teachers and schools to student learning. The calculation of individual level growth is essentially a byproduct, required to estimate teacher and school effects.[6] It is possible to produce student level growth information from VAMs, derived from the residual difference between the predicted and actual student score. However, there is no simple, clear interpretation for this value. VAMs often, as statistician David Betebenner (2011:2) states, "ignore a

[5] See Morgan and Winship (2007) for further discussion.

[6] It is possible to produce student level growth information from VAMs in the form of the residual difference between the predicted and actual student score. However, these are commonly not reported since these may be due to a wide variety of factors not under study.

fundamental interest of stakeholders regarding student growth: how much growth did a student make?"

Betebenner's criticism is not to suggest that developing an understanding of the range of factors that may account for differences in student growth is not important. Rather, it highlights the fact that appropriately characterizing student growth is a necessary first step, independent of attributing cause and effect, to assessing the effectiveness of the educational system. Thus, while VAMs do address some of the limitations of using absolute standards or simple growth scores to assess educational effectiveness, ultimately, they do not provide a measure of student growth that is easily interpretable at the individual student level as well as other critical levels of analysis.

Student Growth Percentiles

Responding in part to criticism of VAM as well as other growth models, Betebenner developed the SGP approach to measuring student learning. The focus of SGP is to understand how much growth a student made compared to students with a history of similar performance ("academic peers"). In this sense, the interpretation of growth is *normative* – how did a student's growth compare to those most like her or him?

Betebenner uses the example of pediatric growth charts to illustrate this approach. Parents are typically less interested in knowing the amount their child grew than in knowing how their child's growth compares to children of a similar age – *their percentile rank as compared to their peers*. Using this approach, the context for interpreting the information is embedded in the measurement itself.

The SGP approach uses quantile regression to estimate the functional relationship between students' prior scores and their current year scores. A series of regressions are run to establish the percentile cut points for each academic peer group. The estimation technique does not assume that the relationship between prior and future performance is linear; instead it relates present achievement to prior achievement with the sum of seven piecewise cubic polynomial functions. This type of specification is common in computational mathematics and is generally intended to approximate a wide variety of different curves (e.g., Burden and Faires, 2001). Ordinarily it is necessary for the researcher to make assumptions about the shape of the curve *a priori*; this approach removes subjectivity by allowing the form of the growth curve to be obtained directly from the data.

SGP produces percentile scores for individual students representing their performance relative to academic peers. While the interpretation is fairly simple – *how much did a student grow compared to the growth of academically similar students* – there are several key aspects of the score that are important to consider.

First, at the individual student level an SGP score has a consistent meaning. A student could be at the 1st percentile or the 99th percentile or anywhere in between, as compared to his peers, just as would be interpreted in a child's pediatric growth chart or a high schooler's SAT score. In contrast, while a student level measure of growth can be extracted from VAMs, it is functionally a standardized residual with little 'real world' interpretation.

Second, students can be combined together to create any group of interest (grade in school, ethnicity, etc.) simply by taking the median of their collective SGP scores. The interpretation of the statistic remains the same – at the group level, the typical (i.e., median*)* student grew at a rate greater than x% of her or his peers. In contrast, VAMs can only provide comparative information about the growth attributable to different groups pre-specified in the statistical model (e.g., those with different teachers or; schools). This is a fundamental difference between the SGP and VAM approaches to growth. While VAM is designed to test differences in growth between groups of students (by teacher, school, etc.) these groups must be specified in advance, and the method is not easily adaptable to describing groups that may not generalize across all students (e.g., students enrolled in a specific program).

Third, the SGP score has low correlation with student demographic characteristics (e.g., income level). The extent to which these characteristics might impact current year performance, and by extension growth, has already been largely accounted for in the model through their impact on prior year performance. As a consequence, when comparing the median SGP across groups of students – *by teacher, program, campus, student demographic group, etc.* – the composition of the groups can be limited as a potential confounding explanation for differences in growth. As a result, educators are able to focus on policy and practice explanations that can be more directly addressed.

3D Growth as a Tool for Improving Student Learning

> *"**3D Growth** elevated the discussion in our district beyond passing rates, to focus on how well are students are learning and growing. If we are going to push our students to compete academically with students globally, we can't be afraid of data showing where we need improvement. Without it, how will we be able to challenge ourselves to become better as a district?"*
>
> *- District Superintendent*

At the core of E3 Alliance's work is a desire to help educators improve the effectiveness of teaching and learning by changing policy and practice. At the request of our partner superintendents, in 2010 E3 spent a year developing criteria, testing models, and adapting what we believe to be the most robust student academic growth modeling methodology in the country to Texas student data. E3 Alliance embraced SGP as an important tool because of the simplicity, flexibility and power it affords measuring student growth at all levels. However, simply providing growth scores to a district without a framework to help teachers, principals, and administrators use them effectively would limit their potential to improve learning.[7] To provide that support, E3 Alliance has integrated growth scores into a set of comprehensive professional development workshops and data visualization tools called **3D Growth**.[8]

Measuring 3D Growth in Texas

A first step in the development of **3D Growth** was adapting the underlying methodology to Texas student data sets. E3 Alliance works closely with The University of Texas at Austin's Education Research Center to access individual level performance data from students enrolled in public schools across the state. This data is used to create **3D Growth** scores for the great majority of students using STAAR data from current and one to three prior test administrations.[9] The only exceptions are those students with no prior test history (i.e., students new to Texas public education) or in the earliest grades (PK-3) or final grades (11th and 12th) where common assessments are not given. Growth data are produced for mathematics and reading STAAR and EOC assessments that have annual measures which can be compared longitudinally. The use of Spanish language and other test types (e.g., Alternative 2)[10] is accounted for in the methodology and does not affect the comparability of **3D Growth** scores across student populations.

It is important to note that a student's **3D Growth** score is based on a comparison to academic peers from across the entire state of Texas, not just those within a single district or region. While 'internal' (i.e., within district) comparisons of learning are important, they provide limited information for benchmarking. This is especially important at the campus and district levels where regional and statewide comparisons may yield important information about best practices. In addition to student level **3D Growth** scores,

---

[7] Several states use SGP scores as an aspect of their accountability framework with a primary focus at the teacher and school. Examples of states incorporating SGP into the accountability frameworks include *Arizona*, *Colorado*, *Delaware*, *Hawaii*, *Massachusetts*, *Michigan*, *New Jersey*, *Utah*, *Wisconsin*, and *Wyoming*.

[8] The "3D" in 3D Growth reflects the Data Driven Decision-making focus of the approach to improve student learning.

[9] E3 uses anonymized 'look-up tables' generated through the Education Research Center to match **3D Growth** scores to individual students using testing histories provided by the districts. The scores are based on the statewide model, but because of FERPA considerations, districts must provide the testing history of their students to E3 Alliance to allow the matching to occur. E3 has developed tools to assist districts with capturing and sharing the necessary data in a secure and efficient manner.

[10] Discontinued test types such as STAAR M and STAAR L are taken into account when looking at prior years(s) STAAR results.

E3 Alliance provides teacher and campus level summaries as well as median growth by student demographic groups (low income, different ethnic groups, ELLs, etc.).

Once all data have been analyzed, E3 Alliance has developed powerful methods for viewing data graphically and multi-dimensionally, showing the performance of individual campuses in terms of **3D Growth** and overall STAAR performance.[11] Graphic representations are also provided to be able to readily identify relative achievement gaps between different student groups on a campus, e.g. positive or negative growth gaps between low income and non-low income students.

Workshop 1 - Measuring 3D Growth and Launching an Action Plan

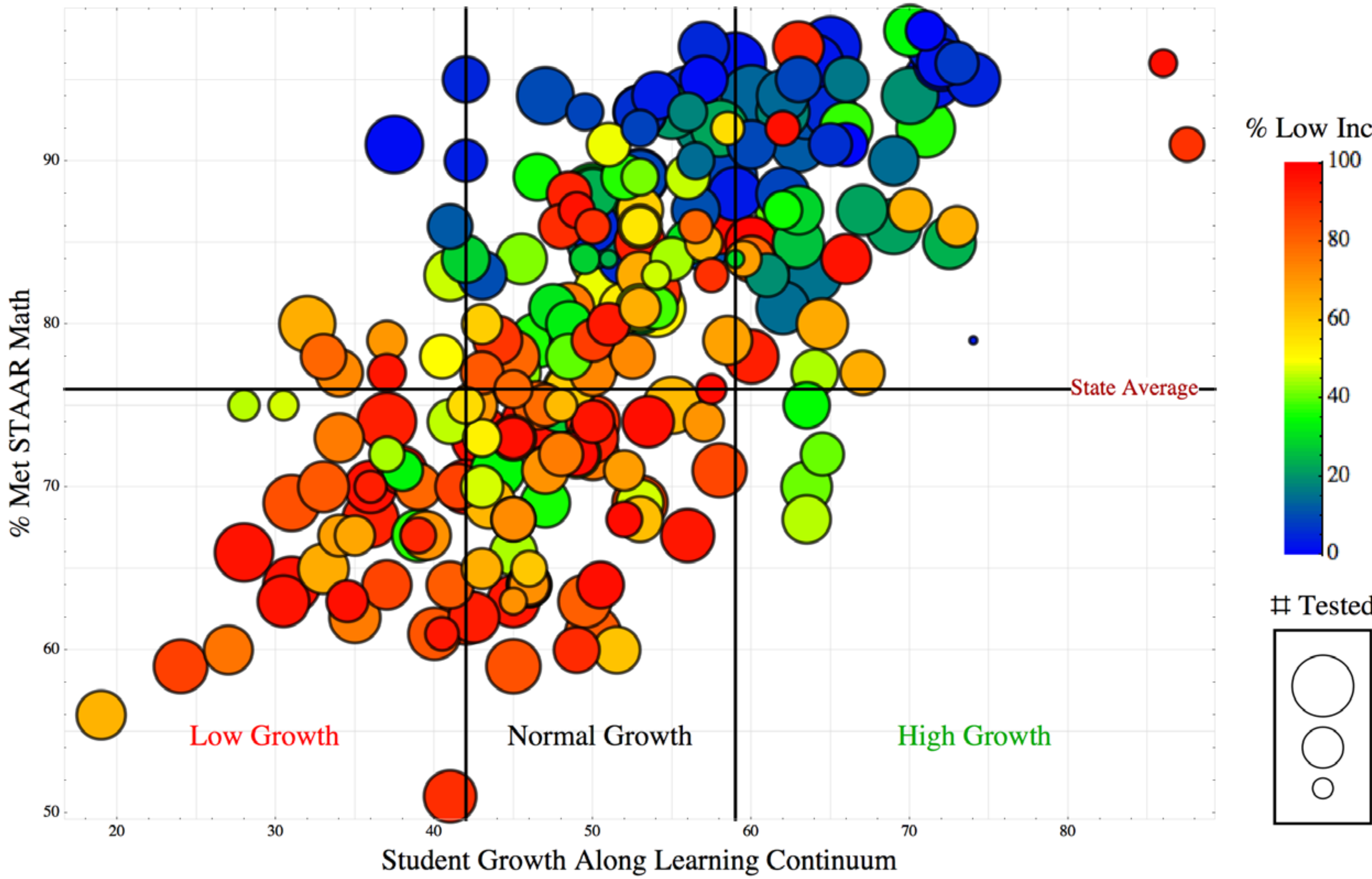
To support districts in understanding **3D Growth** data and learning how to use it to drive school improvement, E3 Alliance typically provides two intensive professional development sessions to school and district administration leaders.

The goals of the first workshop are fourfold. First, participants are introduced to **3D Growth** scores and how they differ from yet complement test score data (i.e., STAAR). Second, participants learn how to interpret **3D Growth** at the campus - comparing campuses across the region by **3D Growth** and STAAR performance – and sub-campus (e.g., by grade level; subpopulation) levels to begin benchmarking performance and looking for areas of strength and need within each school. Third, participants compare multi-year growth data to changes on each campus over multiple years to identify factors that may have positively or negatively contributed to student growth, and use this data to set priorities for improvement in the coming year. Fourth, participants actively engage in process mapping, a powerful tool for building a campus-specific theory of change, and develop a data driven implementation plan for targeting a high-impact area of need to improve student performance.

A fundamental premise underlying **3D Growth** workshops is that viewing data in isolation can provide illumination, but little insight in how to actually improve educational effectiveness. **3D Growth** data jump-starts the improvement process by adding a crucial second dimension to understanding student performance beyond tradition standards models– how much did students grow relative to similarly prepared students? This information supports a more nuanced, and thus actionable, description of learning on a campus. Engagement in the first workshop teaches educators to use their own insights about campus policies and practices to interpret this data to decide *where* to focus to have the greatest impact and *how* to undertake positive change.

[11] Note, these charts label the campuses belonging to that particular district but not those from other districts. To support the identification and adoption of best practices, E3 provides a list of 'Bright Spots' campuses by region exhibiting high levels of growth relative to their similar peers.

Sample 3D Growth Visualization of Regional Elementary School Level Growth and STAAR Performance.



Workshop 2 - Measuring 3D Growth at a Student and Teacher Level

The second workshop expands participants' understanding of how to work with individual level student data while also engaging them in a discussion of teacher level **3D Growth**. Core to both activities is considering both growth and overall performance simultaneously.

At the student level, participants use a Response to Intervention (RTI) framework to develop an understanding of the needs of a variety of students with differing **3D Growth** and STAAR performance profiles. Cross-sectional as well as trend data are integrated with other relevant data to develop model intervention plans for students in need (whether high or low performers) as well as to recognize when an otherwise underperforming student may be making excellent growth that should be reinforced and supported.

At the teacher level, the workshop participants view **3D Growth** summary data for individual teachers alongside STAAR performance information. In contrast to VAM models, which attempt to attribute all outcome differences at the teacher level to the performance of individual teachers, **3D Growth** takes a broader approach. While teacher level **3D Growth** summaries can be used to evaluate teacher performance and are in a number of states, the approach taught in the workshop is that a range of factors – some amenable to measurement and some not – may account for differences in teacher level outcomes. An advantage of **3D Growth** as a measurement tool is that it removes prior performance and student demographics as an explanation for potential differences but leaves it to those who are most familiar with a teacher and campus to understand what is driving performance, whether strong, weak, or somewhere in the middle. As such, **3D Growth** offers a strong and easily interpretable measurement framework for teacher coaching and support that broadens the discussion beyond STAAR performance rates and differences in student populations to how best to increase growth for ***all*** students.

Other Applications of 3D Growth

3D Growth is a powerful tool for understanding and addressing performance improvement planning for individual students, teachers, focused populations (such as 7th grade math students or English Language Learners), campuses and entire districts. **3D Growth** scores can also provide a strong foundation for program evaluation. Aggregating the scores of participants and non-participants in a particular intervention or support program (e.g. dual language instruction; after-school learning supports) provides a simple yet powerful tool for comparing program outcomes that minimizes concerns about background differences confounding program effects. And, because all of the 'behind the scenes' work that happens to calculate and explain **3D Growth** scores, teachers and administrators with varying levels of statistical skill can confidently work with the data to assess the effectiveness of an intervention at the individual student, group, or campus level with confidence.

Conclusion

> *"It is really just as bad technique to make a measurement more accurately than is necessary as it is to make it not accurately enough."*
>
> - Arthur David Ritchie, Chemist and Philosopher

We began by discussing both the ubiquity of standards-based performance metrics in education as well as their limitations. While providing an important measure of overall student achievement, this type of measurement helps little in understanding how much growth is occurring at the student, teacher, campus, or district levels. Without understanding growth, educators are limited in their ability to drive improvements in policy and practice that will lead to higher levels of achievement for all students.

In discussing how best to measure student growth, we considered the potential as well as the limitations offered by VAMs, which have become popular in education for understanding growth at the teacher and campus level. While an improvement on simple gain-score based approaches, VAMs have critical limitations. VAMs are designed to produce particular estimates of how teachers and campuses vary in their contribution to student growth. This focus, with the necessity to control away other contributing factors, comes at the expense of both interpretability as well as generalizability. VAM estimates at the student level have no easily interpretable explanation, represented as a position in a distribution of residuals, and cannot be easily aggregated. At the teacher and campus levels, the outcomes of VAM models are estimates of relative effectiveness rather than clear descriptions of how much learning occurred. Further, VAMs fail to take into account the potentially nonlinear nature of growth, introducing bias into teacher and campus level estimates that may be even stronger at the individual student level.

As Betebenner notes, the use of VAM's "reflects a mismatch between questions of interest and the statistical model employed to answer those questions" (2011:2) – highly specific statistical precision at the expense of interpretability, usability, and generalizability. SGPs embedded within the broader **3D Growth** framework focus on clearly describing how much growth occurred at the student, teacher, student group and campus levels and beyond. Understanding this powerful data is a necessary first step before using the local knowledge of campus and district leaders to interpret the meaning of student growth within the overall context of achievement data and to *use the data to drive toward positive action to improve student outcomes.*

Based on a sophisticated and widely accepted statistical framework, SGP – embedded in the **3D Growth** framework and workshops – is a practitioner-focused tool aimed at engaging educators in an iterative cycle of instructional, programmatic, and policy improvement. The management guru Peter Drucker is quoted as saying "If you can't measure it, you can't improve it". To which we add: *If you can't understand what you've measured, have you really measured anything*?

References

Betebenner, D. W. (2011). *A Technical Overview of the Student Growth Percentile Methodology: Student Growth Percentiles and Percentile Growth Projections/Trajectories*. The National Center for the Improvement of Educational Assessment Dover, New Hampshire

Burden, R. L., & Faires, J. D. (2001). *Numerical analysis* (7th). Prindle Weber and Schmidt, Boston.

Castellano, K. E., and Ho, A. D. (2013). *A Practitioner's Guide to Growth Models*. Council of Chief State School Officers.

Dee, T. S., & Brian Jacob, B., 2011. The impact of no Child Left Behind on student achievement, *Journal of Policy Analysis and Management*, 418-446, Summer.

Morgan, S. L., & Winship, C. (2007). *Counterfactuals and causal inference.* Cambridge University Press.

Neal, D., & Schanzenbach, D. W. (2010). Left behind by design: Proficiency counts and test-based accountability. *The Review of Economics and Statistics, 92*(2), 263-283.

Texas Education Agency (2017). *Calculating the Texas STAAR Progress Measure*. Retrieved from:http://tea.texas.gov/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=25769811320&libID=25769811337.

United States Department of Education (2017). *ESSA Fact Sheet*. Retrieved from: https://www2.ed.gov/policy/elsec/leg/essa/essaassessmentfactsheet1207.pdf

For additional information, contact us at:

E3 Alliance
5930 Middle Fiskville Road, Suite 507
Austin, Texas 78752
512.223.7241 / info@E3Alliance.org

E3 Alliance is a nonprofit 501-(c)(3) organization based in Central Texas providing proven, data-driven services to improve student outcomes across the State of Texas.