

Understanding Value-Added Analysis

What is value-added analysis?

Value-added analysis measures a teacher's effectiveness on a group of students' academic growth from year to year. It uses a student's own academic performance as a basis for determining his or her academic growth and is not related to a student's socio-economic status or other personal characteristics that typically confound achievement-based measures. To dampen the error of measurement from any one single test, value-added analysis uses all student test data simultaneously (STAAR, TAKS, Stanford, and Aprenda from the past three years) within the calculation. HISD uses a value-added system called EVAAS which was developed by Dr. William L. Sanders. Dr. Sanders is a senior research fellow with the University of North Carolina system and is a senior manager of value-added assessment and research for SAS Institute Inc., in Cary, North Carolina. For more information, please see Value-Added and the New STAAR Testing [here](#).

How was the TAKS to STAAR Transition handled?

- Like converting Stanford and Aprenda scores to a common NCE scale with the TAKS to measure growth on two different tests, STAAR scores are also converted to NCE scores. NCE scores for TAKS now represent each student's place in all of the 2011 TAKS scores and for STAAR, represent their place in all of the 2012 STAAR scores. For example, if a student's 2011 TAKS math score is equal to an NCE score of 50 and her 2012 STAAR math score is equal to an NCE score of 50, each score is in the middle of all other scores for that test in each year.
- The Mean NCE Gain is still compared to a growth standard of 0: maintaining the same relative position in statewide student achievement within the year the student tested from one year to the next. The student who had an NCE of 50 in both years made the average amount of growth as the rest of the state in order to stay at the 50th NCE.
- TEA confirmed "one-to-one mapping", meaning EVAAS could use STAAR raw scores even though there were no scale scores.

How has the rigor in STAAR impacted EVAAS results?

The "stretch" of the test, that is, its ability to measure student achievement at the highest levels, is higher for STAAR. While the TAKS historically had enough stretch to perform reliable analysis, EVAAS had revealed that in the final year of TAKS, the district had adapted so well to TAKS, that there was a "ceiling" effect that was preventing EVAAS from measuring growth at the highest levels of achievement in certain subjects. This is common with testing, and the STAAR's greater stretch will allow EVAAS to continue measuring the growth of HISD students with reliability.

Does value-added take into consideration the circumstances of individual students, such as socio-economic or other demographic factors of a school's student population?

At the student level, by including all of a student's testing history, each student serves as his or her own "control" in value-added analysis. This means that value-added measures a student's progress over time and compares a student's performance to their own prior performance and accounts for extenuating circumstances that might affect measurement from any one single test. Value-added uses all student test data simultaneously within the calculation. This minimizes measurement error. So to the extent that socio-economic status or demographic influences persist over time, they are already represented in the student's data used to measure the growth in student academic achievement from one year to the next. This negates the need for adjustment for these influences, and this has been confirmed through research by other leading experts in the field who are independent from EVAAS researchers. The expectation for two

students with the same academic history is that they can make the same amount of growth, regardless of socioeconomic status.

How long has HISD been using value-added data?

HISD has been using value-added data since 2007 and has been providing ongoing professional development and support in the use of value-added information in their school improvement efforts to evaluate program quality and inform instruction. In addition, since 2007, the district has been using campus- and teacher-level value added data to award performance bonuses to teachers, administrators, and other campus-based staff through the ASPIRE Award program. The HISD Board of Education has approved including value-added as one of 34 performance measures in its teacher evaluation system.

Does HISD use data from the first or second test administration?

HISD uses data from a student's first test administration for value-added analysis. Using the student's first test administration provides a fairer comparison to ensure a more valid value-added measure. Using the second test administration would create statistical errors. Additionally, many students who do not score well on the first test administration receive intensive intervention between the first and second test administration. This would create an unfair advantage and bias within the data.

Can you give me more information as to how my value-added scores are calculated?

Yes. To learn more about how value-added analysis is calculated using the SAS EVAAS methodology, download one of the following publications:

<http://www.sas.com/govedu/edu/wrightsandersrivers.pdf>

<http://www.sas.com/govedu/edu/sanderssaxtonhorn.pdf>

For a description tailored to a lay audience:

<http://www.sas.com/resources/asset/SAS-EVAAS-Statistical-Models.pdf>

What is the minimum number of students used to generate a Teacher Value-Added Report?

The ASPIRE Award Program Advisory Committee recommended that more teachers be able to benefit from a value-added report if possible, particularly those who teach in smaller classrooms covering multiple grade levels or in Special Education inclusion settings. SAS EVAAS concurred that its analysis demonstrates that teachers with the equivalent of more than six full-time students can receive reliable reports for this purpose. The HISD Board of Education approved seven as the minimum number of students required in 2008–2009, and this standard is being retained for the new EOC analysis as well.

Why can't you recalculate my value-added score once my reports or my award notice are posted?

SAS EVAAS uses all of the data simultaneously in a complicated data array to construct a multivariate response model (MRM) for grades 3 – 8 or a univariate response model (URM) for End-Of-Course (EOC) Exams. MRM is a layered multivariate longitudinal linear mixed model that produces an estimate of value-added growth that minimizes selection bias and errors associated with measurement. Like the MRM, the URM uses all available test data, but is not longitudinal. These analyses employ sophisticated software and many layers of calculations. Once completed, any re-analysis can only occur at the system level. For this reason, the district conducts a special analysis to ensure that no one is excluded from his or her eligible strands of the ASPIRE Award Program, if data are not available for the teacher at the time of analysis.

Which EVAAS reports are public?

On the evening of October 11, the district made available to the public all district and campus value-added reports with scatter plots, but not lower levels, like individual teachers and students, or diagnostic reports which are for instructional use.

How do you get access to value-added data?

District and campus value-added reports are available to the public through the ASPIRE portal. Access to student value-added data for campus educators is determined by each campus principal.

Which HISD teachers receive individual value-added data?

Teacher-level value-added reports are produced for grades 3–8 in the core foundation subjects of math, reading, and language and also in science and social studies for grades 4–8 for teachers who have at least seven students effectively linked for instruction. In 2012, teachers of courses with EOCs taken by first-time 9th graders also began to receive value-added reports, although theirs are a little different and are discussed in the section called Understanding Value-Added Analysis: End-of-Course (EOC) Exams.

What has research shown about the impact an individual teacher can have on student achievement?

There is indisputable research that shows that teachers matter most when it comes to student achievement. Specific research has been conducted to better understand exactly how much impact teachers can have on student achievement.

Some key findings are as follows:

- Students who had very effective teachers for three years in a row were able to improve their performance on standardized tests by more than 50 percent in comparison to students who had three ineffective teachers in row (Sanders and Rivers, 1996). A similar study conducted in Dallas ISD using a different student assessment and value-added methodology found similar results.
- A teacher's impact on student learning lasts up to four years (Sanders, 2005).
- As teacher effectiveness levels increase, lower-achieving students are the first to benefit (Sanders and Rivers, 1996).
- If a student has an ineffective teacher for two years, this decrease in progress cannot be made up (Rivers, 1999).

Understanding Value-Added Analysis on End-Of-Year Exams: Third through Eighth Grade, Except Courses with End-of-Course or EOC Exams

Why is our campus performance on EVAAS so different than last year? How could our campus change from green to red when there were no dramatic differences in instructional practice?

Since 2007, Houston has used a "Base Year" approach to measuring growth with respect to the State distribution in 2006. Due to the transition from TAKS to STAAR and desiring to have an approach that would be more similar to how the new analysis is measuring growth on EOCs, HISD moved to an "Intra-year" approach in 2012. Houston district growth will now be measured with respect to the state in the current year rather than to a previous year (i.e. 2006). For years, Houston has been increasing annually in growth compared to the state in 2006, so since this is a different growth standard, we expected to see changes in the new reports and far more red across the district. Transitioning to a new, more rigorous test with more stretch is also challenging and different in and of itself, and some of the changes are due to this alone, even if the district had not had to update to a new growth standard.

How is value-added calculated if a student's testing history is incomplete?

The value-added methodology uses up to 5 years of a student's testing history to calculate value-added scores. Students who have no standardized test scores for the year in which the analysis is conducted are not included in a teacher's value-added calculation for that year. In the event that a student has an incomplete testing history, such that the student missed a standardized test in a subject or subjects over the entire testing history, the value-added estimates are calculated based on each student's available testing history across all grades and subjects, making use of the relationship between all subjects and grades for all students in HISD.

How can we understand the differences we noticed in 7th and 8th results, especially math?

There were differences in the test population in these grades in math in 2012. Many accelerated 7th grade students actually tested at the 8th grade level, essentially making them function in the analysis as 8th grade math students with no prior consecutive history in math, which means they carried less weight in the analysis of 8th grade math while the 7th grade analysis lost a significant number of high achieving students from the analysis. At the same time, many high-achieving 8th graders were also removed from the analysis at 8th grade as they switched to Algebra I and took the End-of-Course exams instead of the 8th grade End-of-Year math test.

What will happen to the reports of teachers who taught 7th and 8th grade students who tested above grade level?

At the campus level, high numbers of these students led to no value-added report being possible for some campuses in 7th grade math. While the students who are included in these reports are still in the 8th grade report, teachers of only these students would not have received value-added reports either using our standard value-added analysis (MRM) for grades 3-8 math, because their classes do not have the requisite number of students with at least one prior consecutive math test score. The district asked SAS EVAAS to calculate the value-added results for these teachers using a different methodology (URM) which is the same one now used for End-of Course exams. The univariate response model (URM) does not require consecutive grade level testing. So where a regular report could not be calculated for a teacher who had at least seven tested students at a given grade subject due to off-grade test administration, they will receive this other kind of analysis instead, for just that grade and subject. Their ASPIRE Award will be based on a combination of all their grades in that subject as usual, but it will now be composited across analysis methodologies in addition to being cumulative for all of their grades taught.

What about the students who are still taking TAKS in 10th and 11th grade?

Schools will continue to receive TAKS reports similar to prior years. Mean NCE gain is color-coded based on how that grade level performed relative to the growth standard in that subject. The growth standard remains the 2006 State of Texas Base Year.

2012 Grades and Subjects for TAKS are now:
Reading, ELA, Math – grades 10 and 11
Science and Social Studies – grade 11 only

Which students are included in the School or Campus Value-Added Reports?

All students who tested on a campus in the current year are used to calculate the value-added estimates for the campus. School Diagnostic Reports are based on a subset of the students used in the campus value-added estimates. The students included in the Diagnostic Report must have a test score for this year and for the prior year.

What does it mean when a teacher has a value-added teacher gain index that is negative versus a gain index that is -1 or lower?

When a teacher has a teacher gain index that is between -1 and 1, this means that the progress made by the teacher's students was not detectably different (NDD) from the district as a whole. When a teacher has a teacher gain index that is less than -1, it means that, on average, students taught by the teacher made decidedly less progress than the district.

How are value-added scores adjusted for students testing in Spanish? For example, a student tested in Spanish in 3rd and 4th grades and then tested in English in 5th grade.

All Spanish scores are "mapped" to an English scale. To make these adjustments, SAS EVAAS uses as much student data and information as possible, including Stanford and TAKS, to create relationships between all students' previous and current test scores. By doing this, the information tells us what students would have likely scored had they taken the test in English. The scales for the "Spanish to English" students can then be adjusted so that their progress is equivalent to that of the "English to English" students.

Is it possible for teachers to make positive impact/academic growth with all groups of students (e.g. gifted, special education, low achieving)?

Yes, it is possible for teachers to facilitate high levels of value-added progress with all groups of students. The value-added methodology used is sensitive to individual student's achievement levels. It measures student growth from the end of one year to the end of the next year, regardless of whether a student performs below or above grade level.

How can SAS® EVAAS® use such completely different tests in the same analysis?

State test data from TAKS and STAAR, Stanford, and Aprenda scores are converted to Normal Curve Equivalents (NCE) and anchored to the current state test distribution for each year. This enables value-added scores to be tied to a stable and consistent scale. For more information on how this NCE scale is constructed since STAAR implementation, please see Value-Added and the New STAAR Testing [here](#).

What is an NCE? What is its relative advantage?

NCE stands for Normal Curve Equivalent and is an equal-interval standard score scale with the mean anchored to the 50th percentile. The scale represents a normal distribution of student performance. The advantages of this interval scale are: 1) It can be used to combine the results from multiple tests and to conduct statistical analysis; 2) The size of the NCE unit is sufficiently small to provide a discriminate level of measuring growth; and 3) The scale can be constructed so as to be able to encompass student growth at the highest levels of performance, commonly referred to as having enough "stretch".

I thought all NCEs were the same and you could compare any NCE to another. How are the NCEs used in the SAS EVAAS model different?

NCEs are a statistical construct that have meaning only to the population for which they were constructed. Starting in 2012, SAS EVAAS maps the district's annual state test data and Stanford distributions onto the current state distribution as a standard.

What's the difference between a mean NCE gain score and a gain index, also called the "Gain Score" on the award notice?

An NCE Gain is a conservative estimate of students' academic progress. It is expressed in state NCEs using 2005–2006 as the base year. The standard error provides the basis for establishing a confidence band around the NCE Gain value. One standard error is used in the statistical test reported in comparisons to the HISD Reference Gain. A positive gain indicates HISD made more progress than the state average and a negative gain indicates HISD made less progress than the state average. To calculate a gain index from an NCE gain, first subtract the HISD Reference Gain for that subject and grade(s) from the NCE Gain. Then this difference is divided by the standard error for the NCE gain. The gain index is used by HISD for determining ASPIRE Awards.

How are changes in the timing of the administration of the standardized tests handled?

HISD provides SAS EVAAS (the company that performs the statistical analysis) with the state test data, Stanford, and Aprenda scale scores every year. In creating the new NCE distribution, the scores are statistically adjusted to account for the earlier administration based on the previous year's relationship between the Stanford and state test data.

Why can't kindergarten, 1st, and 2nd grades receive value-added reports?

Value-added analysis is not available in kindergarten and first grade because the testing in these grades does not meet the minimum requirements for use in value-added analysis. Second grade analysis is not available because no reliable baseline measures are available in kindergarten and 1st grade.

Why can't I get value-added reports for my students in 3rd grade Science and Social Studies when we test them on the Stanford/Aprenda environment test in 1st and 2nd grades?

Environment test results do not correlate well enough to the Stanford Social Studies and Science tests administered separately in later grades to be able to use these results in the analysis.

Why does my Teacher Reflections Report show fewer students than I verified (i.e. why are some student results not included in my Teacher Value Added Report)?

A teacher's estimated mean gains include all of the students for whom the teacher claimed at least 20 percent of his/her instructional time during the verification process and who have a test record in that subject this year. However, the students included in the diagnostic report must have a test score for this year in the subject measured and for the prior year in the same subject. If a student does not have a test score for this year and in the prior year in the subject measured, his/her results are not shown in the Teacher Reflections (diagnostic) section of the Teacher Value Added report. The Teacher Reflections (diagnostic) section would likely be a subset of the total number of students used in the teacher's value-added analysis.

How can you say that the value-added score is a reasonable estimate of a teacher's effectiveness when the standard error was so high (e.g. 8.9) on my Teacher Value-Added Report?

Standard error is a measure of uncertainty. Larger standard errors mean that there is more uncertainty about your influence on your students' academic progress. Standard error is influenced by variability in your students' performance and by the number of students in your analysis. If all other variables are the same, teachers with fewer students will have larger standard errors. Additionally, claiming a small percentage of instructional time for a student group effectively reduces the number of students in the analysis. For example, a Science Specialist who claimed 10 percent of the instruction for a group of 60 students has a "class" with an effective size of six students.

If two schools have the same data, but one school has 50 students and one has 100 students, what is the difference in the mean NCE?

The size of the school has little impact on the mean NCE, although it will impact teacher gain indexes. However, the number of fractured/broken student records used in the analysis does impact the mean NCE. More missing student information increases the uncertainty around each estimate, contributing to a higher amount of uncertainty overall for the teacher or building.

How reliable are Student Projection Reports?

Every test score has error of measurement. That's why SAS EVAAS uses all past and current test information from a student to reduce the bias in prior test scores. SAS EVAAS and other independent researchers have determined that a minimum of three prior student test scores are needed to produce a reliable projection assuming the student receives the "average HISD instructional experience". By using as much information as possible about a student, HISD educators receive reliable information with which they can plan appropriate student interventions.

Understanding Value-Added Analysis: Introducing End-of-Course (EOC) Exams

Because there is no consecutive grade testing by subject, like there is for the End-Of-Year exams for 3rd–8th grade STAAR, SAS EVAAS is using a different model for EOCs. EVAAS has previously used this model for analysis of exams like the SAT, for example. It is different in some key ways. There is no District End-of-Course report, because all growth is relative to the district itself. The EVAAS value-added score, called the School Effect, is the Observed Score minus the Predicted Score divided by the Standard Error. The growth standard for campuses is the district average.

End-of-Course Reports look different than **End-of-Year Reports**. Each subject is still reported separately. The information provided includes the number of students used in the analysis, the mean student score, mean student predicted score, the school or teacher effect and how the school or teacher effect compares to the district average by one of the five color-coded categories.

The **End-of-Course Report** scores are reported in actual test scale units, not Normal Curve Equivalent (NCE) units. Because these are calculated and reported in the scale of the specific test they are measuring, the results look different between subjects and different from the **End-of-Year Reports**.

How do I interpret the campus and teacher reports?

The gain index is still interpreted according to the five color system:

Dark green - 2 or more standard errors above the district reference growth

Light green - from 1 to not yet 2 standard errors above the district reference growth

Yellow - growth that is not detectably different from the district reference growth

Light red - from -1 to not yet -2 standard errors below the district reference growth

Dark red - 2 or more standard errors below the district reference growth

Which EOCs received value-added analysis for 2012?

English Reading, English Writing, Algebra I, Biology, and World Geography.

These are essentially the same subjects that are taken by a majority of first-time 9th graders, even though students may be taking courses and testing on some of the other EOCs, such as Geometry. Because this type of value-added analysis is making comparisons within the district, however, the test needs to be administered in a standard manner across the district before EVAAS will consider it to be valid for value-added measurement. As the remaining EOCs are rolled out districtwide for their courses, EVAAS will add them to the analyses.

What's the minimum number of prior test scores a student must have to be included in the EOC analysis?

A student must have at least three prior test scores from any grade and subject to be included.

I completed the district's training on value-added analysis of EOCs, but I can't remember what the predicted and observed scores and the teacher effect are. Can you remind me?

The predicted score is what the analysis predicts that a group of students will receive based on all of their previous testing history (e.g., all the Stanford and TAKS tests in any subject that students in your 2011–2012 Algebra I classes took in the years before they came to you). The Observed or Actual score is the average of just the scores on the specific EOC test that the analysis is measuring (e.g. the actual scores received by your Algebra I students on the first administration of their EOC in spring 2012.) The Teacher Effect is the difference between the two.

But how can you predict what my students will score on an EOC when they've never taken that subject before? Reading predicting math just doesn't make sense to me.

Just as it does in earlier grades, SAS EVAAS exploits the relationships that exist between all tests in order to make a robust prediction of student growth expectations. Yes, there is a relationship between reading and math and between all of the other subjects. These relationships will be different strengths, for example, there is likely to be a much stronger relationship between Algebra I and 8th grade math than 7th grade Stanford Social Studies, but the relationship is still there and is used by the analysis to make the best prediction of student expected growth.

I thought you said the Teacher Effect is the Observed or Actual score minus the Predicted Score, but that's not the number SAS EVAAS put on my report. Why is it different?

Just as they do for the EOY analyses, SAS EVAAS also uses shrinkage estimation to protect schools and teachers from the effects of outliers and other anomalies on the tests. You wouldn't want this year's performance to be compared to a student's "lucky bubbling" from the previous year, any more than you would want a student's bad experience right before testing day for your class to influence your growth directly. In this protective stage of the analysis, students' scores are assumed to be closer to the middle until the evidence pulls them away, either above or below average.

Understanding the Difference between Student Progress and Achievement

What is the difference between student achievement and growth/progress?

A student's actual test score is a measure of student achievement. Often passing percentages are used as a benchmark to show how many students met the minimum standard based upon their actual test scores. Comparing this year's students' passing rates to last year's students' passing rates is a measure of student achievement, not a measure of student growth. Student growth or progress is estimated by comparing a student group's current performance (NCE current year) to their performance last year (updated mean NCE prior year).

Click [here](#) for more information on the difference between achievement and progress.

I went from 50 percent passing to 80 percent passing. How can you say that my students didn't make academic growth?

Passing percentages are an indication of how many students met the minimum standard. Comparing this year's students' passing rates to last year's students' passing rates is not a measure of student growth. Student growth is estimated by comparing a student group's current performance (mean NCE current year) to their performance last year (updated mean NCE prior year). Value-added analysis goes one step further and compares HISD students' growth to a growth standard based on the average growth of similar students across the state of Texas.

My students earn a perfect score every year. How can you show gain for me?

The proportion of students who actually obtain a perfect score two years in a row on the state test is minimal. It is important to understand that value-added analysis is not calculated at the individual student level due to the large amount of measurement error associated with any one student's test score. Value-added analysis compares a group of students' estimated average normal curve equivalent (NCE) this year to their estimated average NCE from the prior year. Including all student data helps to reduce any test's errors of measurement and to ensure all students can be included in the analysis.

Using the Results of Value-Added Analysis

How can value-added information help educators improve teaching and learning?

In combination with formative diagnostic data, value-added information allows educators to assess their impact on student learning and can help initiate conversations about the efficacy of curriculum, instructional practices and programs. Value-added information also allows educators to better identify what is working well and areas for improvement to help individual students and groups of students. Above and beyond the summary information, there is a wealth of diagnostic information being provided through the ASPIRE portal that is appropriate for educators.

There are so many reports. Which reports are most useful as a starting place for conversations with teachers?

Apart from their own individual reports, teachers should focus on the Value-Added Report, Diagnostic, and the Performance Diagnostic for the grades and subjects they teach at their campus, once those reports are released. They also have the ability now to look at individual students so that they can use the Student Projection reports to inform their instruction of their current students. They should refrain from attempting to look at the growth of their own former students as a unit until they receive their teacher reports, however, as these have many critical statistical protections around their individual scores.

What are the advantages of including value-added data as part of the teacher evaluation system?

Nothing matters more to student success than teachers. Having great teachers in the classroom is one of the single-most important things we can do at HISD for meeting the needs of our students and preparing them to graduate on time and be successful in college, careers, and life. Value-added data measure how well schools and teachers are doing in accelerating the academic progress of their students. Value-added data, when used with other measures, provide a more complete picture of performance.

Value-added data, when available, will serve as one measure--not the sole measure--of student performance for both the teacher appraisal and development system and the school leader appraisal for principals. However, it is important to note that the interpretation of the analysis of value-added data for the purposes of the appraisal and development systems is distinct from the interpretation of the data analysis for the purposes of the ASPIRE Award because the specific value-added scores used in each will encompass different numbers of years and grades, and sometimes even subjects, to serve two distinct purposes. The award system sometimes rank orders the scores paying only a certain percent and the appraisal system only looks at a threshold of scores.

If a teacher has a few struggling students in his or her class, will that impact the teacher's value-added results and affect the teacher's job?

The value-added methodology used in HISD to calculate classroom/teacher-level value-added results uses a student's own individual performance as a basis for assessing his or her academic growth from year to year. The contribution that an individual student's data make on the classroom's overall measure of progress is based on the academic progress that each student in the class makes relative to his or her previous academic history.

Conservatism is built into the value-added methodology by using analysis that assumes a teacher is average until the evidence shows otherwise. This protects individuals from receiving a disproportionately extreme value-added score due to random error, especially important for small class sizes. Teachers should focus on adding value, or growth, for all of their students by using the data to tailor instructional strategies to students' needs.

What can I do to improve my value-added scores?

Evidence shows that when teachers and schools collaboratively focus on students as individuals and provide the targeted curriculum and instruction that meet each student's learning needs, student progress accelerates. Measuring intra-year progress and using that information in formative ways to ensure that all

students are making progress is critical. Involving students in the process of setting instructional targets and measuring their growth also has been proven effective. Using value-added and other supportive data sources can help HISD educators target areas for professional growth. And, finally, working collaboratively with colleagues on instructional practices and engaging targeted professional development are essential. HISD staff members can engage in learning paths through ASPIRE•Learn to guide their professional development as well.

How do I best use student projections—especially if students have a low probability of passing the test?

Student projections are a tool to help ensure that students are placed appropriately in classes or into instructional groups. These projections also help educators apply the necessary resources to meet all students' learning needs.

To learn more about HISD's 2011 value-added reporting, send an e-mail with your specific questions to aspireaward@houstonisd.org.