Understanding Value-Added Analysis

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.

How can SAS® EVAAS® use two completely different tests in the same analysis?
TAKS, Stanford, and Aprenda scores are converted to Normal Curve Equivalents (NCE) and anchored to the 2005–2006 Texas TAKS distribution. This enables value-added scores to be tied to a stable and consistent scale.

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. HISD maps our annual TAKS and Stanford distributions onto the Texas 2005–2006 TAKS 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.

Why is the TAKS Accommodated now included?
HISD first had three years of TAKS Accommodated data for all subjects in 2009. Therefore, TAKS Accommodated results are now reliable enough to be included for value-added analysis.

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.

**How was the early administration of the Stanford 10 and Aprenda 3 handled?**
HISD provides SAS EVAAS (the company that performs the statistical analysis) with the Stanford and Aprenda scale scores. In creating the new NCE distribution, the Stanford and Aprenda scores are statistically adjusted to account for the earlier administration based on the previous year's relationship between the Stanford and TAKS test data.

**How was the change in the Stanford 10 norms from 2002 to 2007 handled?**
Because the full range of NCE values are recalculated every year by SAS EVAAS to align with the 2005–2006 base year, there was no need to first convert Stanford scale scores between years as the norm changed from 2002 to 2007. This shift in norms is automatically embedded in the NCE values utilized for all reports.

**How are value-added scores adjusted for students testing in Spanish?** For example, a student tested in Spanish in 3rd and 4th grades, then tested in English in 5th and 6th grades.
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.

**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 can't we get individual Teacher Value-Added Reports at the high school level?**
High school students have varied schedules and often take classes out of sequence or repeat courses. The test that they are required to take is based on what grade level their campus has them placed in. This can lead to disconnects between courses (and consequently teachers) and high school TAKS outcomes. For this reason, high-school level value-added has been calculated at the department and grade level only. When end-of-course tests replace TAKS tests, individual teacher value-added reports for high school courses will be calculated.

**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.

**Why does my Teacher Value-Added 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.

**Why was the estimate on my 2009 value-added report different last year from what I see this year for my 2009 value-added report?**
Each year, HISD’s value-added information becomes more robust as more longitudinal information is included in the value-added analysis. Calibrations are made to historical value-added information to ensure a fair comparison from one year to the next. Consequently, it is possible that current value-added reports may reflect a different result for last year than was reported in last year’s analysis. ASPIRE Awards are based on the most reliable data available at the time the awards are calculated.

**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:

For a description tailored to a lay audience:

**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). 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. 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.

**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 the Difference between Student Progress and Achievement

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.

We went from being an ‘Acceptable’ to a ‘Recognized’ campus. How can you say progress didn’t occur? Why didn’t we earn bonus pay in Strand I of the ASPIRE Award Program?
Strand I of the ASPIRE Award Program is based on the composite of all students’ academic growth and in all core subjects using the TAKS or the Stanford/Aprenda in grades in which TAKS is not given. Campuses are rank ordered into quartiles based on the composite value-added score for all subjects tested at that campus compared to that of other campuses. Only those ranking in the top half qualify a staff member to earn an award for this strand. Conversely, accountability ratings such as ‘Acceptable’ or ‘Recognized’ are based on the percentage of students who meet a minimum achievement standard. Moving from one campus rating last year to a higher rating this year is an indication that this year, more of your students met that minimum achievement standard than last years’ group of students. This outcome is not necessarily an indication that most of your students made high levels of academic progress. Additionally, accountability ratings are based solely on TAKS tests. Therefore, a campus might be rated highly for meeting a student achievement level based on passing rates, but not show enough growth in value-added scores for all its students to be ranked in the top half of all HISD campuses.

Does the TAKS have enough “stretch” to be used for value-added analysis?
A test has enough stretch when there are enough difficult questions to discriminate achievement between high-performing students. However, there will always be some students who make perfect scores on a test. This does not mean that there is an issue with “stretch” or there is a “ceiling effect” at the teacher-, department- or campus-level. After looking at the distribution of TAKS data over several years, HISD and SAS EVAAS have determined that there is no concern about “stretch” with the TAKS.

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 TAKS 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

There are dramatic drops in many subjects for the 6th and 10th grade value-added scores, particularly at the district and school levels. How do you explain these differences?

These drops have been persistent every year, but were particularly strong in 2008–2009. SAS EVAAS conducted an analysis of this phenomenon, in order to isolate whether it is endemic to the tests given at those grade levels, or to other factors. SAS EVAAS concluded that because the drops occur across most, but not all schools, and because they are evident for other tested subjects, that the drops are due to factors other than the change in tests in those grades (going from TAKS to Stanford). Possible explanatory factors include the variability by which curricula is taught in certain grades and subjects, the sudden breakdown in curricula alignment from the prior grade in certain subjects, and to adjustments that occur in these grades when more students change environments and as a group may perform more poorly on tests than expected.

How can I show high gains for last year, but not this year, when I did nothing different?

Value-added analysis estimates the influence the curriculum and instruction had on a specific student group during a given year. If your results change from one year to the next with a different student group, consider these questions: 1) Could the curriculum and instruction have been better suited to the needs of one student group than another? 2) Did any other changes on the campus influence students’ academic growth (e.g. schedule, instructional support, intervention, student population, etc.)?

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 2010 value-added reporting, send an e-mail with your specific questions to aspireaward@houstonisd.org.