Offer Name: Interactive Achievement | SAS
Name of Assessment: Interactive Achievement Student Growth Assessments

Interactive Achievement/SAS	**1 - Interactive Achievement Student Growth Assessments**	**Math, English, Science, History**	Are the items that are included on your premade tests excluded from the item bank used by teachers? Could you also clarify procedures to validate measures of growth and the training protocol?

DOE Question #1: *Are the items that are included on your premade tests excluded from the item bank used by teachers?*

IA/SAS Response: Yes, the items included on the premade Interactive Achievement Student Growth Assessments are excluded from the item bank used by teachers.

DOE Question #2: *Could you clarify procedures to validate measures of growth?*

IA/SAS Response:

SAS EVAAS utilizes two methodologies for modeling student growth—MRM and URM (described in our original RFP response). The first approach can be used when there is a consecutive grade given test, and the second can be used for all testing scenarios. They both share some of the same general benefits. Some of those general benefits speak to the rationale of these approaches. In general, each model uses all available testing history to sufficiently dampen the measurement error in test scores, and they do not exclude students with missing data. Below is the specific rationale for each measure of student growth methodology.

Advantages of the SAS EVAAS MRM Approach:
There are numerous advantages to the SAS EVAAS MRM value-added modeling, some of which are listed below:

- All students are included in the analyses, even if they have missing test scores. Each student's entire or complete testing history is included without imputing test scores.
- By including all students in the analyses, even those with a sporadic testing history, provides the most realistic estimate of achievement available for a district or school.
- It minimizes the influence of measurement error by using up to five years of data for an individual student. Analyzing all subjects simultaneously increases the precision of the estimates.
- Because the influence of measurement error is minimized, there is no need to adjust the estimates for socio-economic factors.
- It allows educators to benefit from all tests, even when tests are on differing scales.

Advantages of the SAS EVAAS URM Approach:

There are many advantages to the SAS EVAAS URM approach, some of which are listed below:

- It does not require students to have all predictors or the same set of predictors, as long as a student has at least three prior test scores in any grade/subject.

- It minimizes the influence of measurement error by using all available data for an individual student. Analyzing all subjects simultaneously increases the precision of the estimates.
- Because the influence of measurement error is minimized, there is no need to adjust the estimates for socio-economic factors.
- It allows educators to benefit from all tests, even when tests are on differing scales.
- By using shrinkage estimation, it protects the effects from misclassification as effective or ineffective.
- Accommodates the teaching scenarios where more than one instructor has responsibility for a student's instruction in a particular grade/subject, such as team teaching.

Specific to the teacher model, there are distinct advantages of the SAS EVAAS approach:

- Allows for comparisons among teachers at different schools, within an entire district, or across the state.
- Allows for the possibility that a teacher may be very effective in one subject yet ineffective in another.
- Accommodates teaching scenarios where more than one instructor has responsibility for a student's instruction in a particular grade/subject, such as team teaching.
- The EVAAS teacher model for tests, such as the state CRT, allows teacher effects to accumulate over time. That is, how well a student does in the current subject/grade/year depends not only on the current teacher, but also on the accumulated knowledge and skills acquired under previous teachers.
- Using shrinkage estimation, the model protects against misclassifying teachers as effective or ineffective, particularly in the case of teachers with few students.

Additional information can be found at http://www.sas.com/resources/asset/SAS-EVAAS-Statistical-Models.pdf

SAS EVAAS provides many procedures to validate data, analyses, and output. Prior to conducting the analysis SAS preps, cleans, merges, and reviews the quality of data to ensure it is appropriate for use in value-added analyses. Validation for our student growth measurements or value-added measures is conducted through internal empirical reliability and other empirical validation that checks other ways of measuring growth and making comparisons. There have also been external validation of the EVAAS models.

Prep, Clean, and Merge Data Prior To Analyses

SAS considers the quality of information to be the most important aspect of a value-added model used to determine educator effectiveness. In fact, reliability of information has been—and always will be—the primary driver behind the development of all SAS value-added and projection models. SAS has extensive experience, gleaned from four statewide implementations, in the cleaning and merging of student records over time, allowing us to uphold high standards of data cleanliness from year to year. Virginia Department of Education (VA DOE) will benefit from SAS's data processing and quality control methodology.

SAS will receive student data VA DOE schools each year with the appropriate code definitions and file formats and host the solution to provide the services specified in this RFP. SAS checks student test data and demographic data for consistency each year before incorporating it into a database that links students over time. File formats are compared annually to assess any changes from one year to the next. SAS will also receive teacher data, program data, and student-teacher linkages from the VA DOE schools each year. These linkages will be analyzed to determine match rates to students. SAS will work with the VA DOE or the testing vendor to provide applicable reporting. SAS will deliver reports identifying any data inconsistencies or omissions. These reports proactively

allow the VA DOE to validate the input data. In past experiences, SAS has been able to catch issues in the coding of indicator variables that were not discovered with the testing vendor until after the files were sent. Other errors in student testing data have also been found where test vendors did not resubmit corrected files to the client, and SAS was able to detect these errors. It is of the utmost importance that every part of information is checked for errors to ensure usable and clean data.

SAS understands that every education partner has different needs when analyzing a variety of data. With that in mind, the following procedures are intended to give an idea of the work SAS does to maintain quality of data before the analyses are run. While the below list is not comprehensive, it provides an outline of common SAS data preparation procedures. If SAS determines that other data may be needed for the appropriate value-added modeling to be conducted, then SAS will provide the VA DOE with a detailed list of these data and determine what may be used moving forward to aid in a successful implementation.

Merging Records Over Time

SAS will conduct stringent data management to produce value-added estimates and individual student projections by using multiple element merge keys. Using this technique, student and teacher records are matched over time. For students, the multiple element merge key could consist of the state student identification number, student first name, student last name, student birthdate, student middle initial, and student district number. For teachers, the multiple element merge key could consist of the new teacher licensure number, teacher first name, teacher last name, and district where he or she teaches. These elements are used to merge student records and link teachers accurately and reliably.

Checking Statewide Distributions of Scale Scores

Analyses can only be as good as the data that is used. Furthermore, models will behave differently depending upon variations in the statewide student testing scores.

It is important to examine the statewide distribution of student scale scores each year to determine if they are appropriate to use in a longitudinally-linked analysis. It is also important to ensure the appropriate value-added model does not make assumptions not met by these scales. Typically, scales must meet the following three requirements to be used in most types of analysis:

1. Scales must have sufficient stretch to show measurable differences among students at the very top and bottom of the distributions.
2. Scales must be reliable from year to year.
3. The test must be highly correlated with the curricular objectives.

SAS checks the first two requirements every year using the statewide distribution of scale scores along with the sample of student test data that is sent each year before the full test data is given. SAS will aid the VA DOE in checking these distributions and provide reporting back to the VA DOE around the statewide distribution of scale scores each year. These checks are vital to providing a valid and reliable value-added analysis. If the distributions do not meet the second requirement, then SAS may alter models to eliminate any effect of year- to-year change in test scales. When necessary, SAS offers options to alleviate the concerns on the first requirement. In general, some approaches exist that remove the differences of scales from year to year. As shown below, some value-added models, such as predictive-type models do not depend on test scaling. In fact, different tests can be used in different years with different scales to provide the most comprehensive use of data.

Most state and district administrators design or select tests that assess what students are expected to learn during their standard course of study; therefore, the third requirement is typically met in practice. However, if the third requirement is not met, then all districts, schools, and/or teachers might not be measurably different in terms of value-added measures. SAS monitors this aspect of measurable differences annually.

Data Quality of Student Test Scores Prior to the Analysis

SAS's extensive experience handling the below data quality issues has proven to increase the accuracy of the SAS EVAAS solution. SAS will collaborate with the VA DOE to determine how to best utilize pre-processing procedures based on existing and future policy decisions. For an example of some of the business rules used to ensure data quality, please see the list below.

- **Missing Grades:** If a grade is missing on any End-of-Grade tests, then these records will be excluded from the analysis. If a grade is missing on any End-of-Course tests, then the record will take on the most likely grade based on End-of-Grade testing or, alternatively, the grade in which that test is offered. Policy decisions can alter the modification of grades in End-of-Course tests for the VA DOE.
- **Duplicate Scores:** If a student has a duplicate test score for a particular subject and grade in a given testing semester, then the extra score(s) will be excluded from the analysis and reporting.
- **Missing Some District or School Records:** If a student has a missing district or school for a particular subject and grade in a given testing semester, then the record that has a district and/or school will be included over the record that has the missing data.
- **Multiple Student Records in the Same Semester:** If a student has multiple records in the same semester for a particular subject and grade and those tests are given on different days, then the first administration is used. If there is clearly a first administration and second administration for retesting, then only the first administration should be used. If a student has multiple records in the same semester for a particular subject and grade and those tests are given on the same day, and the test scores are not the same, then those scores will be excluded from the analysis. Again, these can be altered by policy decisions, but they are the recommended practices to ensure an equitable comparison when providing value-added analyses.
- **Multiple Grades in Same Subject and Year:** Typically, a student should not have different grades in the same subject in the same year. If that is the case, then the student's records are checked to see if the data for two separate students were inadvertently linked. If this is the case, then the student data is adjusted so that each unique student is associated with only the appropriate scores. If the scores appear to all be associated with a single unique student, then scores that appear inconsistent are excluded from the analysis.
- **Unexpected Grade Changes:** If a student skips more than one grade (e.g., moves from 6th in 2009 to 9th in 2010) or is moved back by one grade or more (i.e. moves from 4th is 2009 to 3rd in 2010) in the same subject, then the student's records are examined to determine whether two separate students were inadvertently linked. If this is the case, then the student data is adjusted so that each unique student is associated with only the appropriate scores. If the scores appear to all be associated with a single unique student, then scores that appear inconsistent are excluded from the analysis.
- **Multiple Schools in the Same Semester:** If a student tested at two different schools in a given semester, then the student's records are examined to determine whether two separate students were inadvertently linked. If this is the case, then the student data is adjusted so that each unique student is associated with only the appropriate scores.
- **Outliers:** Student assessment scores are checked each year to determine if they are outliers when considering all of the other scores in a reference group of scores from the individual student. These reference scores are weighted differently depending on proximity in time to the score in question. Scores are checked for outliers using related subjects as the reference group. For example, when searching for outliers for math test scores, all math subjects, both End-of-Grade and End-of-Course type tests, are examined simultaneously, and any scores which appear inconsistent, given the other scores for the student, are flagged. Scores are flagged in a conservative way to avoid excluding any student scores that

should not be excluded. Scores can be flagged as either high or low outliers. Once discovered, that outlier will not be used in the analysis, but it will be displayed on the student testing history. This is done as part of a data quality procedure to ensure no erroneous scores are used.

Empirical Stability

Because SAS EVAAS modeling approaches use up to five years of student testing history, even the first year of implementation can provide multi-year value-added estimates to districts, schools and teachers. Based on the SAS EVAAS value-added estimates provided to another client, the stability of our teacher model can be assessed by how teachers' value-added estimates have changed over time. While some variation can be expected (teachers may improve in effectiveness over time, they may teach different grades or subjects, and have different students every year, etc.), high repeatability of individual teacher's estimates can indicate the model's robustness. Our internal research on the SAS EVAAS teacher value-added estimates reported correlation coefficients between 0.70 and 0.80 for three-year averages in different grades and subjects. Furthermore, a teacher who was identified as very effective at the beginning of his or her career was extremely likely to remain effective and above average several years later (about 65 – 70% remained effective). Similarly, a teacher who was identified as very ineffective at the beginning of his or her career was fairly likely to remain ineffective several years later (about half remained ineffective).

The stability of our teacher estimates can be seen another way. Recent research by the SAS EVAAS team found that when teachers move to a new school or district, they carry most, if not all, of their teaching effectiveness with them. This finding was consistent regardless of teachers moved within the same socioeconomic strata or across strata. See Stability of Teacher Estimates for the full report.

Empirical QA

Every time EVAAS analyses are performed, there are a series of checks that look at the model's results. Some of these checks include comparing EVAAS estimates to other ways of measuring growth to see if there are any major differences. If there are, then the data is examined to determine how to explain those differences.

Each year, in some models new data is used to re-estimate prior year results based on the most complete set of student testing. These estimates are compared to their original estimates from prior years. If any of the estimates differ by a certain amount, then those are examined to determine why we are seeing a difference. Estimates are compared over time as discussed in the previous section. The value-added measures from each model are also compared to other measures of student characteristics across a state to see if there are any strong relationships. If those relationships do exist, then that is examined further.

Efficacy

SAS EVAAS's customer success stories are a testament to our long term efficacy. Many of our customer success stories can be found through the following link SAS EVAAS Customer Success Stories.

Third Party Validation

SAS EVAAS reporting has passed rigorous scrutiny by several federal agencies. More specifically, the US Government Accounting Office has reviewed SAS's projection source code for accuracy and four US Department of Education Peer Review Committees have approved SAS reporting for use in growth waivers in NCLB.

- As reported in **Ohio's growth model application for NCLB**, with four prior scores, the multiple correlation of SAS EVAAS individual student projections is higher three years in advance than the simple correlation

between adjacent years.[1] In other words, EVAAS projections to an eighth grade math test for current fifth grade students are more closely correlated to students' actual performance than their seventh grade math scores.

- The **State of Tennessee** recently published empirical data regarding its projections, and, regardless of the projected subject and grade, at least 90% of the students who had at least a 70% probability of success were in fact either proficient or advanced on their tests (tvaas.sas.com).

Additionally, the following research from the RAND Corporation corroborates the stability and reliability of EVAAS modeling approaches.

- On the **choice of a complex value added model**, please consult: McCaffrey, D. F., Han, B., and Lockwood, J. R. (2008). "Value-Added Models: Analytic Issues." A paper prepared for the National Research Council and the National Academy of Education, Board on Testing and Accountability Workshop on Value-Added Modeling, November 13 and 14, 2008, Washington D.C.
- On the **advantages of the longitudinal, mixed model approach**, please consult: Lockwood, J. R. and McCaffrey, D. F. (2007). "Controlling for Individual Heterogeneity in Longitudinal Models, with Applications to Student Achievement." *Electronic Journal of Statistics*, Vol. 1, 223-252.
- On the **insufficiency of simple value added models**, please consult: McCaffrey, D. F., Han, B., and Lockwood, J. R. (2008). "From Data to Bonuses: A Case Study of the Issues Related to Awarding Teachers Pay on the Basis of the Students' Progress." Paper presented at the conference on Performance Incentives: Their Growing Impact on American K-12 Education, Feb. 28-29, National Center on Performance Incentives at Vanderbilt University.

Recently, WestEd, a non-profit education research and development agency, reviewed and evaluated possible models for use in North Carolina's accountability model, and its final recommendation were models provided by SAS. For more information please refer to http://www.ncpublicschools.org/docs/stateboard/highlights/2012/02highlights.pdf

DOE Question #2: *Could you also clarify the training protocol?*

IA/SAS Response:

Training for users (i.e. system administration, curriculum staff, principals and assistant principals, coaches, teachers, etc.) may be completed using a train-the-trainer model, an on-site train-the-user model, and/or through additional online training options.

For the delivery of Interactive Achievement Growth Assessment's, training is focused on the creation of test sessions and the access to both IA and SAS EVAAS reports.

IA training sessions are hands-on and in person, unless the district opts for online training. If in person, more than one training session may be held simultaneously, and multiple sessions may be scheduled per day depending upon the division's preference and IA's availability.

[1] See, for example, the May 1, 2007 Addendum to the Growth Model Proposal available at http://www2.ed.gov/admins/lead/account/growthmodel/oh/index.html.