Rewarding Excellence in Instruction and Leadership

2015-16 Conceptual Guidebook: Components of the REIL Score

A practitioner’s guide to understanding the REIL performance measures.
Contents

Components of the REIL Score ................................................................. 3
How is Observation Used to Measure My Performance? ................. 5
Frequently Asked Questions about the Learning Observation Instrument.............................................................. 8
How is Student Growth Used to Measure My Performance? ........ 10
Frequently Asked Questions about Student Growth Measures........... 15
Key Terms in the Guidebook................................................................. 18
Components of the REIL Score

Components of the REIL Performance Evaluation System

As shown in Figure 1, the REIL performance evaluation system conceptually has three major components: Classroom Observation (50 percent), Individual Growth (40 percent), and School Growth (10 percent).

In general, a teacher’s evaluation results within each component are converted to performance scores on a Common 1 to 5 Scale. Each component’s performance scores are then combined to calculate a final REIL Score and an aligned performance classification.

Classroom Observation

The Learning Observation Instrument (LOI) supports ongoing feedback to teachers on their instructional practice over the course of a school year. Teachers undergo multiple observation cycles during the year, each consisting of a pre-conference, classroom observation, and post-conference. For each cycle, teachers are rated on 22 elements of teaching based on a 0-5 scale. Cycle scores are aggregated into a final observation score that is converted to the Common 1 to 5 Scale.

In addition to calculating the total LOI points across all cycles, observation scores are also used to create a measure based on consistency of proficient scores called the LOI 3+ Rate. For instance, a teacher who receives ratings of 3s on all elements will have a higher LOI 3+ Rate than one who receives 1s on some elements and 5s on others, even though the two teachers may earn the same total number of LOI points. The LOI 3+ rate is designed to reward teachers who consistently demonstrate proficiency across the LOI elements.
Measures of Student Growth

There are two general approaches to measuring student growth. They are: Value-Added Modelling (VAM) and Categorical Growth Analysis (CGA). Both approaches are based on comparing students’ actual academic growth to their expected academic growth.

Student Growth: Value Added Modeling

Value-added modeling (VAM) is used to estimate the contribution of an individual teacher or school to student learning over the course of the school year. These contributions are measured using students’ test scores on the AzMERIT, Galileo, and Custom Assessments.

Value-added modeling uses statistical techniques to compare students’ actual test scores to their predicted test scores. For each subject area, predictions are made about what a student is expected to achieve on a particular assessment based on what is typical for students who are in the same grade, have similar prior test scores and demographic characteristics, and are enrolled in a school with similar demographic characteristics. Students’ actual test scores are then compared to their predicted test scores in order to determine their value-added achievement gain in the subject area. If students exceed expectations, the gain is positive. If they fall short of expectations, the gain is negative.

Student Growth: Categorical Growth Analysis

In addition to value-added analysis, categorical growth analysis (CGA) is used to evaluate student growth. The CGA model used in the REIL Score is calculated by Assessment Technology Incorporated (ATI). CGA compares observed growth to expected growth for students over a defined period of time. A statistical method called the repeated measures t-test is used to determine whether observed growth is significantly different from expected growth. Based on the results of this test, any given group of students is then categorized by the extent to which those students, on average, outperformed, met, or underperformed their growth expectations.
How is Observation Used to Measure My Performance?

Classroom observation results account for 50 percent of the REIL Score. This section provides an overview of the observation tool and process used by REIL Alliance districts to measure the quality of teachers’ instructional practice. It also describes how the results of observations will be used to inform a teacher’s REIL Score. More detailed information on the REIL observation tools and processes for all educators is available on the REIL Website.

The Learning Observation Instrument (LOI) and Observation Cycle

Teachers in REIL Alliance districts participate in multiple LOI Observation Cycles during a school year, enabling ongoing conversation about instruction among teachers, administrators, instructional coaches, and REIL Peer Evaluators. As seen in Figure 2, each LOI observation cycle consists of a pre-conference, classroom observation, and post-conference.

Figure 2. The LOI Observation Cycle

Instructional practices are assessed during this observation process using the Learning Observation Instrument (LOI). The LOI is the result of a collaborative effort to create a cross-district observational tool to be used as part of a performance-based evaluation system. It is designed to define effective teaching practices, encourage dialogue about instruction, and support differentiated professional growth. The LOI is comprised of six rubrics: Content, Formative Assessment, Instructional Strategies, Learner Engagement, Learning Community, and Professional Responsibilities. The LOI is aligned to the Model Core teaching standards.
All rubrics, except Professional Responsibilities, are implemented uniformly across REIL Alliance districts. Professional Responsibilities is scored at the discretion of each district and not used in the REIL Score.

Each LOI rubric contains multiple elements resulting in a total of 22 unique elements, as shown in Table 1. Each element is rated either during the pre-conference, classroom observation, or post-conference phase of each observation cycle, with one exception. Task Analysis, in the Content rubric, is rated during both the pre-conference and classroom observation. Thus, during each observation cycle, a total of 22 elements are rated, the majority of which are assessed during the classroom observation. Evaluators rate each element on a scale from 0 to 5 according to the criteria provided in the rubrics.

Table 1. LOI Rubrics and Elements

<table>
<thead>
<tr>
<th>Rubric</th>
<th>Element</th>
<th>Setting</th>
<th>Pre- Conference</th>
<th>Classroom Observation</th>
<th>Post-Conference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content</td>
<td>Conceptual Understanding</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Task Analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Connections to Content</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Content Accessibility</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formative Assessment</td>
<td>Real-Time Assessment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Student Progress</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Correct level of Difficulty</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructional Strategies</td>
<td>Teacher Role</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Instructional Approach</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Practice/Aligned Activity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Feedback</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Monitor and Adjust</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Analysis of Instruction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learner Engagement</td>
<td>Student-to-Student Interaction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Teacher-to-Student Interaction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Authentic Engagement</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Critical Thinking</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learning Community</td>
<td>Routines and Procedures</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Responsibility for Learning</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Monitoring Student Behavior</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Relationships</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Element Ratings per Cycle:</td>
<td></td>
<td></td>
<td>4</td>
<td>16</td>
<td>2</td>
</tr>
</tbody>
</table>
Focus of the Pre-Conference

The pre-conference is an opportunity to discuss the lesson that will be observed as well as the lesson-planning process. The evaluator may ask questions for clarification, review any data and artifacts presented by the teacher, and document the discussion. During this time, a teacher will be scored on the four pre-conference elements of Task Analysis, Content Accessibility, Real-Time Assessment, and Correct Level of Difficulty.

Focus of the Classroom Observation

The observation is an opportunity for a teacher to present the complete lesson discussed in the pre-conference. The evaluator will observe the lesson within five school days of the pre-conference. During the classroom observation, the evaluator will script the lesson as well as students’ responses and actions. Each lesson will be scored on 16 observation elements of the LOI.

Focus of the Post-Conference

The post-conference gives a teacher an opportunity to analyze and reflect on his/her lesson as well as discuss students’ assessment results. The evaluator may ask questions for clarification, review any data and artifacts presented by the teacher, and document the discussion. The evaluator will score a teacher on the two post-conference elements of Student Progress and Analysis of Instruction.
Frequently Asked Questions about the Learning Observation Instrument

Why are multiple observation cycles required per school year?

Research shows that in order to reliably measure the quality of your instruction, the teaching and learning in your classroom must be observed on multiple occasions.¹ A single lesson provides just a snapshot of your instructional practice. Taking multiple snapshots will yield a more complete picture of the teaching and learning that occurs in your classroom, and in turn provide you with more reliable feedback on how to improve your instruction. All teachers know that some lessons are more successful than others. Each additional observation reduces the risk that your overall learning observation results for the school year will be affected by a single lesson that does not go as planned.

What happens if I am not able to complete my district’s minimum number of observation cycles? Will I still be eligible for a REIL performance award?

There may be some circumstances that preclude a teacher from being observed in his/her district’s minimum number of observation cycles (e.g., late hire, short-term leave, etc.). In these cases, eligibility is determined per district discretion.

How can I monitor my Learning Observation results throughout the school year?

Teachers have the ability to access, review, and verify their scores online through the REILize Decision Support System (RDSS).

What should I do if I notice a reporting error in my LOI results?

Evaluators are responsible for entering the results of your observation cycle into RDSS: Profile. If you notice an error in your LOI results, you should discuss the error with your evaluator.

**What steps are taken to ensure that all evaluators rate the LOI rubric criteria with the same levels of rigor and objectivity?**

It is essential for all evaluators to use the same standards when rating the LOI criteria. To ensure this is the case, all evaluators receive qualified and certified evaluator training. In order to establish and maintain inter-rater agreement, evaluators participate in co-observations, and an annual certified evaluator assessment.

**Why doesn’t the professional responsibilities rubric count toward my REIL Score?**

The REIL Score is currently informed by elements that can be observed through classroom observation. The professional responsibilities rubric cannot be scored during a classroom lesson. In addition, each district handles professional responsibilities outside of the formal evaluation process according to its established policies and procedures.
How is Student Growth Used to Measure My Performance?

This section provides a conceptual explanation of the measures used to assess teachers’ impact on student growth.

What does it mean to measure student growth?

Whether student growth is estimated using Value-Added Modeling (VAM) or Categorical Growth Analysis (CGA), student growth refers to the difference between a given student’s actual test score and an expected test score based on what is typical for students with similar pre-test scores.

Student growth differs from student achievement. Student achievement provides information, at one point in time, about whether a student has mastered academic content. In contrast, student growth looks at how student performance changes from one point in time to another.

Teachers are evaluated on both their individual students’ growth as well as the growth of all students in the school in which they teach. The individual growth provides a measure of how a teacher is performing with his/her individual students. Teachers are also measured at the school level, because teachers are part of a school-level team with the shared goal to improve the performance of all students from one year to the next.

Measures of Student Growth

For REIL Scores, there are two general approaches to measuring student growth. They are: Value-Added Modelling (VAM) and Categorical Growth Analysis (CGA). Both approaches are based on comparing students’ actual academic growth to their expected academic growth. There are important distinctions between the two approaches and how each contributes to a teacher’s REIL Score.

In general, the use of VAM and/or CGA depends on the following factors:

- The types of assessments administered by the school at which the teacher serves (e.g., AzMERIT, Galileo, and/or Custom Assessments).
- The types of assessments administered to the students for whom the teacher is responsible (e.g., AzMERIT, Galileo, and/or Custom Assessments).
- The student growth component being measured (i.e., individual or school).
Benefits of Using Value-Added Student Growth Measures

With regards to its use in instructional and educator evaluation systems, the practical benefits of VAM include the following:

- **VAM is the only student growth model that levels the playing field by controlling for factors outside of the educator’s control, such as student background and classroom and school characteristics.** By controlling for other factors that also impact student achievement but are beyond the control of the teacher or school, VAM estimates the unique contributions of a particular teacher or school to student learning over time.

- **Compared to other test-based evaluation models, VAM facilitates apples-to-apples comparisons across teachers or schools by adjusting for pre-existing differences in students’ individual learning abilities and family backgrounds.**

- **VAM captures student growth across the achievement distribution, including the growth of students who perform below state proficiency standards.** Regardless of where students begin the school year academically, educators are evaluated based on their students’ improvement over the course of the year.

---

**Student Characteristics Controlled for in the Model**

Value-added models evaluate teacher or school effectiveness by tracking student test score growth between testing periods using statistical techniques accounting for factors that could also impact student learning but are beyond the control of teachers and schools. When appropriate and the requisite data are available, the 2015-16 value-added models will control for the following characteristics at the student, classroom, and school levels:

- Prior achievement
- Gifted and talented status
- Special education status
- Free and reduced-price lunch eligibility status
- English Language Learner (ELL) status
- Fluent English Proficient (FEP) status
- Migrant status
- Homeless status
- Student mobility status
- Student attendance rate in prior academic year

**Assessments Used to Measure Value-Added Student Growth**

The individual and school value-added contributions to student learning are measured using students’ test scores on the Galileo, and Custom Assessments. We refer to these measures as the VAM-Z, VAM-G, and VAM-C respectively.

- **THE VAM-Z measure of student growth** is based on student test scores in ELA and math on the state assessment, AzMERIT.

- **The VAM-G measure of student growth** is based on student test scores in reading, math, science, and writing on the Galileo Assessments.

- **The VAM-C measure of student growth** is based on student test scores on the Custom Assessments in social studies, music, band, choir, physical education, theatre, dance, art, Tech Literacy, and Transition in a range of elementary, middle, and high school grades. MCESA, in collaboration with approximately 40 Local Education Agencies, developed the Custom Assessments for the non-state-tested subject areas to support the collection of valid and reliable assessment data to measure student academic progress.
Benefits of Using Categorical Student Growth Measures

With regard to its use in instructional and educator evaluation systems, the practical benefits of CGA include the following:

- Educators in REIL districts have used the Galileo Assessments and CGA scores for a number of years to assess student growth and inform instructional practices.

- CGA measures are based on student growth rather than on achievement, which provides important signals regarding instructional effectiveness on students across the entire performance distribution.

Assessments Used to Measure Categorical Student Growth

The individual and school categorical growth contributions to student learning are measured using students’ test scores on the Galileo Assessments. We refer to this measure as the CGA-G. The CGA-G measure of student growth is based on student test scores in reading, math, science, and writing on the Galileo Assessments in grades K-12 (subject dependent).

Alternate Measures of Student Growth

Some educators may have special circumstances where it is not possible to conduct VAM or CGA. For example, a teacher may be responsible for a small number of students, or a class may not have an available test that has been vetted as reliable and valid. In those cases,

What Is Categorical Growth Analysis?

*Categorical Growth Analysis considers the relationship between the growth observed for a group of students over a given time period and the growth expected for those students in that time period. Categorical Growth Analysis employs a well-established common statistical test, the repeated measures t-test, to evaluate whether the difference between observed growth and expected growth is significant for the group of students for which a teacher/principal is responsible.*

alternate growth measures may be employed if appropriate data are available and the measures are sufficiently valid and reliable for the purposes of gauging student growth.

Types of measures may include:

- **Piggyback Method**: This approach will be applied to educators who instruct students in subject areas that are not covered by the Custom Assessments or Galileo Assessments. Specifically, the Individual Growth component for the REIL Score will be measured using a VAM student growth model and Galileo Assessments data from the students an educator instructs, but in a subject area different than the subject area taught.

- **Percent Met Growth Method**: This approach will be applied to educators who instruct fewer than 10 students in every subject-area and grade-level combination. For this approach, the Individual Growth component for the REIL Score will be measured by examining the percent of students who met their expected growth in the VAM student growth model.

- **Practice Related Growth Index (PRG)**: This approach will be applied to educators who instruct fewer than 10 eligible students in any and all subject areas. The PRG Index is a measure of practice, based on a subset of elements from the LOI. The PRG Index is not a direct measure of student growth, but it is related to student growth. PRG measure is calculated by summing the total points earned for each identified element score and converting the total points into a performance score (1-5).

The following five LOI elements are included in the PRG Index:

<table>
<thead>
<tr>
<th>Rubric</th>
<th>Element</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formative Assessment</td>
<td>Student Progress</td>
</tr>
<tr>
<td>Instructional Strategies</td>
<td>Practice/Aligned Activity</td>
</tr>
<tr>
<td>Learner Engagement</td>
<td>Teacher-to-Student Interaction</td>
</tr>
<tr>
<td>Learner Engagement</td>
<td>Authentic Engagement</td>
</tr>
<tr>
<td>Learning Community</td>
<td>Routines &amp; Procedures</td>
</tr>
</tbody>
</table>
Frequently Asked Questions about Student Growth Measures

Our student population is highly mobile and students often transfer in and out of the school throughout the year. Will these students be included in our growth results?

For the 2015-16 REIL Score, mobile students may be included in student growth measures if educators, school leaders, and district leaders attribute students through the assessment confirmation process to the Galileo and Custom Assessments. All students included in the assessment confirmation process will be used in the CGA student growth measure using Galileo Assessments, including students who may not have been enrolled a full academic year. The Arizona Department of Education (ADE) will identify students who are enrolled in the school for the full academic year, as defined by enrollment at the school during the first 10 days and enrollment at the school during the final 10 days. The VAM student growth measures may only include students who are identified as full academic year, which in some cases may be a subset of the students included in the CGA student growth measure. The VAM student growth measure will also control for the effects of student mobility on student growth by including a covariate for students who are new to the school at the beginning of the school year owing to non-promotional school transfers. The CGA student growth measure does not include a similar control for student mobility.

---

3 This variable is designed to capture mobility that is not due to normal promotions from elementary to middle school and middle school to high school. Prior studies have found this type of mobility associates with lower academic achievement (Ballou, Sanders, & Wright, 2004).
I have heard of SGP. Is that the same thing as VAM or CGA?

The Student Growth Percentiles (SGP) model is different than the VAM and CGA student growth models used for REIL Scores. The SGP model assigns a percentile rank to students after comparing a student’s actual score to the actual scores of other students with similar past test scores. The student growth percentile ranks for growth are relative to other students during the same time period. In the SGP model, an educator is assigned the median SGP of all the assigned students. Unlike the VAM, but similar to the CGA, the SGP model does not account for student background characteristics.  

I don’t know how to interpret value-added model teacher results.

VAM estimates provide information about how much an educator contributed to student learning in a particular subject-area and in a particular year. To generate teacher results, VAM controls for factors outside of a teacher’s control and then compares actual scores with expected scores for students assigned to the teacher. All VAM growth measures are expressed as standardized scores. These standardized scores benchmark the performance of a teacher’s or school’s students relative to the average level of performance in a given student population, after taking into account their prior test scores and demographic characteristics. Specifically, each growth measure is expressed in Standard Deviations relative to the population average. The teacher effects are standardized using data from the REIL districts. A positive teacher result indicates that the teacher was responsible for more student learning occurring in that classroom than what would be expected to occur in a similar classroom with students of similar background characteristics.

I have heard that teacher results from value-added models are significantly impacted by differences in the students assigned to teachers.

VAM compares each student’s actual test score to an expected test score that is based on what is typical for students with similar pre-test scores, after controlling for student, classroom, and school characteristics. For example, the actual test scores of students who are English Language Learners (ELL) or eligible for free- or reduced-price lunch (FRL) will be compared with the expected test scores for similar students. An educator responsible for students who are lower performing on the pre-test but make academic growth that is larger than expected can generate a larger teacher effect from VAM than an educator responsible for higher performing students who make the academic growth that was expected for them. In this regard, VAM helps level the playing field and enables educators to demonstrate the impact of their instruction on student learning.

Why do different student growth models using the same data produce different results?

The primary reasons why student growth models (VAM, CGA, or SGP) may produce different results using the same data are due to differences in: (1) the students who are linked to educators; (2) the way in which growth expectations are calculated; and (3) how the models control for student, classroom, and school characteristics. However, results from different student growth models using the same data are correlated, which means that for the majority of educators, the results will be comparable.

What causes REIL Scores to vary from year to year?

REIL Scores may vary from year to year because of variation in the components that comprise a REIL Score, as well as variation in the student growth models. Results from both of the student growth models (VAM and CGA) may fluctuate from year to year for some teachers. This fluctuation can be the result of a variety of reasons, including the quality of the proctoring of the student assessment and variation in the performance of student cohorts. However, for the vast majority of educators, teacher results from student growth models are robust from year to year and reflect how much an educator contributed to student learning.

When I look in the Assessment System (e.g., Galileo or PCG) where my student assessment data is displayed, I see data and test results. How does that information connect with the VAM or CGA scores?

The Individual Growth and School Growth components of REIL Scores are based on VAM or CGA that use scale scores in sophisticated statistical calculations. The resultant VAM or CGA estimates are then run through a conversion chart to identify that measure’s REIL Performance Score and corresponding REIL Score points. The Assessment System does not display VAM or CGA calculations or conversions used in the REIL Score. See the REIL Score practical guide for more information about the actual calculations used in REIL Scores.

What is the score that students need to get on the post-test?

VAM and CGA are calculated annually based on how groups of students perform on the pre-test. The expected growth target for a post-test varies for each subgroup and varies each year. Therefore it is not possible to identify a particular score that needs to be achieved.
Key Terms in the Guidebook

**Common 1 to 5 Scale**
The scale on which the results of all performance measures are placed before they are combined to produce the REIL Score. Conversion charts are used to place the Total LOI Points, LOI 3+ Rate, Individual Teacher Growth Measures, Team Growth Measures, and School Growth Measures on the Common 1 to 5 Scale.

**Conversion Charts**
Charts that explain the criteria for translating results from the various LOI and student growth measures to the Common 1 to 5 Scale.

**Galileo Assessments Categorical Growth Measure (CGA-G)**
A growth measure developed by Assessment Technology Incorporated (ATI). This measure tracks whether students, on average, underperformed, met, or outperformed their test score growth expectations based on their mastery of Arizona state content standards from the fall-to-spring administrations of the assessment.

**Galileo Assessments (VAM-G) and Custom Assessment (VAM-C) Value-Added Modeling**
An estimate of the contribution of an individual teacher or school to student growth. A value-added estimate represents the difference between students’ actual and predicted test scores, where the predicted test scores are determined based on students’ prior achievement and demographic characteristics.

**The Learning Observation Instrument (LOI)**
The instrument used to measure the quality of teachers’ instructional practice through classroom observations. The LOI is comprised of six rubrics: Content, Formative Assessment, Instructional Strategies, Learner Engagement, Learning Community, and Professional Responsibilities.

**LOI 3+ Rate**
The number of LOI element ratings that teachers receive from all observation cycles that are rated a 3 or higher on the LOI rubrics.

**LOI Observation Cycle**
The three phases in each evaluation of classroom instruction: Pre-Conference, Observation, and Post-Conference. Teachers are expected to complete multiple observation cycles in 2014-15. A single observation cycle should take ten school days to complete.

**Performance Score**
The 1 to 5 score associated with each of the REIL Scorecard’s performance measures: the Total LOI Points, LOI 3+ Rate, Individual Teacher Growth Measures, and School Growth Measures.
**REIL Components**
The REIL Scorecard’s three categories of performance measures: Classroom Observation, Individual Growth, and School Growth.

**REIL Score**
The measure of educator effectiveness determined by combining a teacher’s performance on the five REIL Components. The REIL Score is then used to assign state performance classifications and REIL Awards. The REIL Score will range from 100 to 500.

**Scale Score**
A conversion of a student’s raw score on an assessment to a common scale that allows for numerical comparison of that student’s score to the scores of other students who took different versions of that test.

**Standard Deviation**
A measure of the extent to which the scores from an assessment vary from the mean score for that assessment. Standard deviations help determine the significance of differences between scores, whether they be students’ test scores on a standardized assessment or educators’ scores on an observational assessment.

**State Performance Classifications**
The four evaluation designations that districts must assign to Arizona classroom teachers: Ineffective, Developing, Effective, and Highly Effective. In the REIL system, the “Effective” category is partitioned into “Effective 1” and “Effective 2” in order to differentiate performance awards.

**Total LOI Points**
The total number of points that teachers accumulate from their evaluators’ ratings of LOI elements across all observation cycles.