**Indicator: Unit pre-tests and post-tests are administered to all students in the grade level and subject covered by the unit of instruction. (92)**

**Explanation:** Teacher Instructional Teams develop units of instruction with one or more formative assessments to determine student mastery of objectives prior to the introduction of lessons or units and their mastery at the end. This is a quick-check that enables the teacher to adjust his or her approach in teaching the lesson/unit and differentiate assignments and supports for each student.

**Questions:** Do your Instructional Teams systematically develop and administer formative assessments at the beginning and end of lessons or units? Do the teachers use the data to differentiate instruction? Do the teams use the information to modify units of instruction and share the most successful teaching strategies?

A unit test is an assessment device, aligned with each standards-based objective covered in the unit, and administered to all students before and after the unit of instruction (or smaller parts of the unit). The pre-test and post-test are the same test, or parallel items for the same objectives, given at the beginning and end of a unit. In some cases, especially in the lower grades, the unit test is divided into a series of smaller tests, given before and after instruction in the objectives covered on the smaller test. Unit tests are constructed to give teachers a good idea of a student’s current level of mastery of the objectives without taking a great amount of time to administer. A unit test need not be a pencil and paper test, especially in the lower grades, but is a way for the teacher to specifically check each student’s mastery of each objective in a manner that is not time consuming (Redding, 2007).

Effect size is calculated by taking the difference in two mean scores and then dividing this figure by the average spread of student scores (i.e. average standard deviation), or: Effect Size (ES) = [Average of the post-test scores – Average of the pre-test scores]/Average standard deviation. To be valid, the spread of scores should be approximately distributed in a ‘normal’ bell curve shape. Effect sizes can be used to: investigate the effectiveness of a particular intervention for a defined group of students; compare the effectiveness of different interventions; or evaluate the growth over time. Some researchers believe it is important to keep in mind such questions as: “How well is what I am doing working for different groups of students each year and why?” “What possible reasons could there be for some student or groups of students progressing more or less?” and “How does student progress compare with their achievement levels?”

Despite Race to the Top and similar legislation, effect size is only a single measure of progress, and educators are encouraged to use a range of learner achievement and multiple measures of data to complement existing achievement measures in order to reliably understand and replicate evidence of what works. Effect size for cohorts smaller than 30 are often not suitable for reliably estimating the impact of an intervention, and Hattie (2012) cautions that care needs to be taken while interpreting any findings for small sample sizes, as outliers in student scores can skew the effect sizes and may require special consideration. Effect sizes derived from small sample sizes and individual student effect sizes should only be used indicatively by the teacher to ask such questions as What possible reasons could there be for why that group of students recorded these estimated effect sizes? What will we do for students who are achieving at expected achievement levels but not the expected growth effect size?

In his 2013 article Art and Science of Teaching: How to Show Student Learning, Marzano said, “Since Race to the Top legislation, teacher evaluation systems across the United States
have emphasized measures of student learning – precisely because Race to the Top requires the inclusion of such measures in a teacher’s evaluation” (p. 82). State departments of education commonly use state test scores to calculate measures of student learning – to accommodate Race to the Top’s requirements. Although these ‘growth scores’ or ‘value-added measures’ have an initial appeal, they’ve been roundly criticized, including for the fact that, when the same group is measured twice using two different tests to compute value-added measures, the results can be widely different. He recommends developing other ways to demonstrate that students taught by a given teacher have learned. He also has a couple of recommendations: (#1) Use common assessments to create common growth measures, and (#2) Use common student surveys to create common growth measures.

**For Special Education**

For students with disabilities, set the stage by first explaining to students the purpose of a pre-assessment (not for a grade, but to find out what they already know and don’t know about the upcoming unit of study so that the teacher can plan instruction accordingly). Then administer the common formative pre-assessment (or individual classroom or program pre-assessment) to the students. Students with disabilities and English language learners (ELLs) often differ from their classmates in the ways they respond to testing/assessment situations, which is why it is important to set the stage (Ainsworth, 2011).

**References and Resources**


©2013 Academic Development Institute