Student Evidence Tracker (SET)

ESSA Level 4

Monitoring and Tracking Real-Time Student Progress in any Learning Environment

1-800-979-3316 | LearningSciences.com
SET: Formative Assessment and Student Academic Achievement – Evidenced Based Theoretical Rationale

“Formative assessment is a process that engages teachers and students in gathering, interpreting, and using evidence about what and how students are learning.”

Research indicates that students who participated in formative assessment perform better than those who do not.

Formative assessment guides educator decisions to make midstream adjustments to instruction that address learner needs in a timely manner. Student Evidence Tracker (SET) immediately boosts student engagement and learning because students self-assess their learning and teachers have minute-to-minute data from which to take actionable feedback prior to testing. SET rises to ESSA Level 4 because it “demonstrates the rationale” that students who self-assess their learning goals are more effective academically.

Long-Term Efficacy Study
Efficacy proven by using randomized control group

Level 1
Strong Evidence of Impact

Level 2
Moderate Evidence of Impact

Level 3
Promising Evidence of Impact

Level 4
Rational for Expecting Impact

Long-Term Efficacy Study
Efficacy proven by comparing results between students or schools

Long-Term Efficacy Study
Efficacy proven by showing correlation with certain kinds of users

Long-Term Efficacy Study
Solution design aligns with learning sciences principles
SET: Supporting Research

Across 19 studies that met WWC house standards without reservations, formative assessment had positive effects on student academic achievement. The average of these effect sizes was 0.26 standard deviation, which is just over the benchmark set by the WWC for a substantively important effect size (greater than 0.25 or less than –0.25). However, the effect sizes ranged from –0.46 to 1.22 with formative assessment in math having the largest effects (0.36) on student academic achievement:

Table 1. Mean effect sizes for formative assessment, by subject area

<table>
<thead>
<tr>
<th>Subject area</th>
<th>Number of studies</th>
<th>Number of effect sizes</th>
<th>Mean effect size</th>
<th>Standard deviation</th>
<th>Minimum effect size</th>
<th>Maximum effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math</td>
<td>6</td>
<td>10</td>
<td><strong>0.36</strong></td>
<td>0.33</td>
<td>-0.18</td>
<td>1.01</td>
</tr>
<tr>
<td>Reading</td>
<td>7</td>
<td>12</td>
<td>0.22</td>
<td>0.45</td>
<td>-0.46</td>
<td>1.22</td>
</tr>
<tr>
<td>Writing</td>
<td>6</td>
<td>8</td>
<td>0.21</td>
<td>0.24</td>
<td>-0.20</td>
<td>0.63</td>
</tr>
<tr>
<td>Spelling</td>
<td>2</td>
<td>4</td>
<td>0.19</td>
<td>0.09</td>
<td>0.09</td>
<td><strong>0.30</strong></td>
</tr>
<tr>
<td>Composition</td>
<td>4</td>
<td>4</td>
<td>0.22</td>
<td>0.35</td>
<td>-0.20</td>
<td><strong>0.63</strong></td>
</tr>
</tbody>
</table>

**Note:** The table presents descriptive statistics for effect sizes across studies. Bolded values indicate effect sizes greater than the What Works Clearinghouse benchmark for a substantively important effect size (greater than 0.25 or less than -0.25; U.S. Department of Education, 2014b). See tables B2 and B3 in appendix B for the statistical significance of the effects in individual studies.

**a.** The column sum (10) does not equal the total number of studies reviewed (22) because 3 studies (Craven, Marsh & Debus, 1991; Fuchs, Butterworth & Fuchs, 1989; and Mostow et al., 2003) did not provide enough information to calculate effect sizes.

**b.** The number of effect sizes is greater than the number of studies because two studies of math (Fuchs, Fuchs, Hamlett, & Steker, 1991; and Ysseldyke & Tardrew, 2007), four studies of reading (Fuchs, Fuchs, Hamlett, & Ferfuson, 1992, Johnson, Graham, & Harris, 1997; Martens, Eckert, & Begeny, 2007; and McCurdy & Shaprio, 1992), and two studies of writing (Fuchs, Fuchs, Hamlett, & Allinder, 1991a, 1991b) included more than one comparison for which effect sizes could be calculated.

**Source:** Authors’ analysis of studies published between 1988 and 2014; see appendix A for details.
Intervention Effects Were Stronger When Formative Assessment was Student-Directed

Within the 19 studies that met WWC house standards without reservations, formative assessment research data was derived from 3 sources: educator-directed, computer program-directed and student-directed. The student-directed formative assessments proved to be the most effective:

Table 3. Mean effect sizes for formative assessment, in math by type

<table>
<thead>
<tr>
<th>Subject area</th>
<th>Number of studies</th>
<th>Number of effect sizes</th>
<th>Mean effect size</th>
<th>Standard deviation</th>
<th>Minimum effect size</th>
<th>Maximum effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student-directed</td>
<td>4</td>
<td>4</td>
<td>0.45</td>
<td>0.49</td>
<td>-0.18</td>
<td>1.01</td>
</tr>
<tr>
<td>Other-directed</td>
<td>3</td>
<td>6</td>
<td>0.30</td>
<td>0.21</td>
<td>0.07</td>
<td>0.66</td>
</tr>
</tbody>
</table>

This is one of the main reasons why LSI developed SET – because it is student-based, self-directed formative assessment. While SET is in the process of being vetted for ESSA Levels 3 and 2, the other iteration of the technology, Standards Tracker®, where teachers track students’ progress to the standard has shown evidence at ESSA Level 3.

The sample study, that was extracted from Standards Tracker, was based on 540 formative and diagnostic assessment scores matched to standards and includes six teachers and 138 students. Six teachers at one elementary school in Florida consistently tracked student performance in the classroom to state standards. Students’ last scores for each standard assessed at the classroom level were used as the final outcome measure. The table on the next page shows the level of association between the two metrics. The correlation coefficient was positive and significant (.357) and demonstrates that both variables are related. This implies that the higher the tracker score, the higher the value of the end of year test score; or it can also mean that when the tracker score is lower, the test score also will be lower. The data, being statistically significant (Sig. 2-tailed < 0.01), also shows that the correlation is not just by chance. Hence, from this test, we can say that there is a moderate association between the two scores.
These results are significant to the field because there are few studies that examine whether student performance in the classroom aligns with more robust yet lagging end of year assessments. Furthermore, the evidence demonstrates ESSA Level 3 evidence on a similar technology than SET. Future research will be critical in establishing evidence levels for student-directed formative assessment in the classroom.

### Correlations Between Formative Tracker and End of Year Diagnostic Scores

<table>
<thead>
<tr>
<th>End of Year</th>
<th>Formative Tracker</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Correlation</td>
<td>.357*</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.000</td>
</tr>
<tr>
<td>N</td>
<td>138</td>
</tr>
</tbody>
</table>


Much of the research used in this report is from The Marzano Center: [https://www.marzanocenter.com/](https://www.marzanocenter.com/)

Our vision for education is to close the achievement gap. Equip all students with the social, emotional, and cognitive skills they need to thrive in the 21st century. Expand equity by giving every child access to rigorous core instruction that empowers learners to free themselves from generational poverty.

©2020 Learning Sciences International. #LSI52-04