Analysis of State Test Data

In Spring 2013, New York State administered the first ELA/Literacy and Mathematics Common Core tests in Grades 3 – 8. These tests were intended to provide students, families, educators, and the public better measures of student proficiency in the knowledge and skills students need to succeed in college and careers. The New York State Grade 3 – 8 Common Core Tests are designed primarily to classify student performance into one of four levels. Appropriate analyses of test results can provide meaningful information to help inform discussions about instructional decisions.

The New York State Education Department (NYSED) strives to support and to promote uses of state test data that lead to valid interpretations about student performance. It is important to understand how to use and interpret test data to ensure appropriate instructional decisions are made. NYSED provides some individual student performance information on a Parent Report. The Parent Report includes student scale scores, performance levels, and subscores. Additionally, student performance on subscores may be aggregated and analyzed at the classroom, school, and district levels by your local data service provider.

This document focuses on possible approaches to analyzing state test data, other than those provided by NYSED in the Parent Report, and guidance for using these analyses to inform instruction, given the type of information available from and provided to districts, RICs, and BOCES. Use caution when conducting some analyses regardless of their source because the results may be misleading. Table 1 provides a summary of recommended and discouraged analyses.

<table>
<thead>
<tr>
<th>Performance Data</th>
<th>Is the Analysis Recommended at this Level?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Student</td>
</tr>
<tr>
<td>ELA: Informational Standards vs. Literary Standards</td>
<td>Y</td>
</tr>
<tr>
<td>Math: Performance on each Major Cluster</td>
<td>Y*</td>
</tr>
<tr>
<td>ELA or Math: Performance grouped by Standard</td>
<td>N</td>
</tr>
<tr>
<td>ELA or Math: Performance on Individual Released Test Questions</td>
<td>N</td>
</tr>
<tr>
<td>ELA or Math: Performance on Individual Non-Released Test Questions</td>
<td>N</td>
</tr>
</tbody>
</table>

Note. Y = yes; WSC = with strong caution; N = no; * = recommended when there are 9 or more questions in a cluster.

1 Descriptions of the four performance levels are provided on score reports and on EngageNY.org (http://www.engageny.org/resource/performance-level-descriptions-for-ela-and-mathematics).

**General Guidance on Evaluating Analyses**

When deciding whether it is appropriate to conduct a specific analysis of test data, it is important to consider two things: (a) the number of questions included in the analysis; and (b) the number of students included in the analysis. The relationship between these two components is presented graphically in Figure 1.

As is shown in Figure 1 below, interpretations should be made with a low level of confidence when reviewing the performance of a single student on a single question (the solid, blue line). More confidence is warranted when the data are aggregated to examine a group of students at the classroom level (the dashed, red line), the school level (the dashed, green line), or the district level (the dashed, purple line), even with a single question. For example, if the majority of students in the school answer a single question incorrectly, one can have confidence that the students in the school generally do not display mastery of the skills and abilities represented by the standard measured. In addition, if a single student performs well on nine or more questions aligned to a single standard, the level of confidence increases with regard to mastery of the knowledge and skills represented by that standard. Figure 1 shows a high degree of confidence when examining performance in relation to a single question at the school or district level.

**Figure 1: Relationship Between Number of Questions and Number of Students Examined in Analysis**

**General Guidelines for Calculating and Interpreting Results**

The greater the number of questions examined for a single student, the greater the confidence in the interpretations about the student’s performance on questions grouped in meaningful ways within ELA or Math (see Table 1). Because of this, NYSED chose not to compute subscores for groups of questions for which the total number of points was fewer than nine. This decision was made to ensure that the
subscores presented on the score report provide reasonably stable evidence to support an interpretation of the student’s performance in a given area. It is strongly recommended that the nine-question minimum be kept in mind when evaluating and determining additional analyses conducted at a local level.

Performance Data Analyses

1. ELA: Informational Reading Standards vs. Literary Reading Standards

The ELA tests contain two types of passages: informational and literary. Districts and schools may choose to analyze student performance on questions grouped by the type of passage with which the questions were associated: literary or informational. Students answer questions related to both types of passages. Your district may choose to create subscores to compare the number of questions students correctly answer related to these two types of passages.

2. Math: Performance on each Major Cluster

Districts and schools may choose to report information about performance in each major math cluster. Cluster performance can be defined as the number of points earned for correctly answered questions measuring the standards in that cluster. Districts and schools may choose to compare performance across major clusters.

3. ELA or Math: Performance grouped by Standard

Districts and schools may choose to report information about performance on individual standards. Performance on individual standards would be calculated by summing the scores on questions answered correctly for all the questions addressing a specific standard (e.g., Grade 3: interpret products of whole numbers). Districts and schools may choose to compare performance across individual standards.

4. ELA or Math: Performance on Individual Released Test Questions

Districts or schools may choose to report information about performance on individual questions. Inferences drawn from results on individual test questions are the least reliable (see Figure 1) and are often inaccurate. Over reliance on inferences made from data at the question level may lead to making shifts in instructional emphasis that are not warranted.

5. ELA or Math: Performance on Individual Non-Released Test Questions

NYSED strongly discourages making inferences – regardless of level of aggregation – from single test questions that were not released on EngageNY.org. For non-released questions, the only information available is the standard or standards which the question assesses. Making an inference about a standard based on a single item without access to that question prompt for context purposes has an even greater likelihood of leading to faulty interpretations about student performance.

---

iii A cluster is defined as a group of related standards.
performance on that standard. Such faulty interpretations often lead to educational practices that do not improve student learning or performance on subsequent tests.

**How Can Schools and Districts Use This Information?**

Schools and Districts can leverage these analyses in a number of ways. Once they have examined the data and identified trends and areas for growth, schools and districts may:

- Cross-reference standards encompassed in the subscore with the curriculum used to teach those standards.
- In identified growth areas, analyze curricular materials using the Tri-State/EQIP Rubric and the Publisher’s Criteria.
- Consider high-quality professional development on the instructional shifts and the use of aligned curriculum.
- Determine priorities for district goals based on identified areas of growth.
- Consider professional development for teachers targeted at the content addressed in standards with weak student performance.