Minnesota Department of Education

How to Use the Mathematics Benchmark Toolkit

The Mathematics Benchmark Toolkit helps educators focus their efforts for improved mathematics achievement. The toolkit provides state-level data about the difficulty levels of the mathematics benchmarks and a suggested process for using the state data along with school and district benchmark reports to help educators identify concerns and areas for improvement.

The state level mathematics benchmark-item difficulty data provides information about the difficulty levels of the benchmarks. It is important to note that these average benchmark-item difficulties reflect *the characteristics of the items in the test bank at a specific point in time, and not the inherent difficulty of the benchmarks themselves*. This data identifies for each grade level (grades 3-8 and high school) the following information: 1) benchmarks whose items on average are most difficult, 2) benchmarks whose items on average are least difficult, and 3) benchmarks whose items on average are of average difficulty for each grade.

To help schools and districts identify possible improvement actions for increased student achievement, the state-level mathematics benchmark-item difficulty data has been added to two different formats of the *2007 Minnesota K-12 Academic Standards in Mathematics*. The two formats offer very different ways to think about the standards and benchmarks in order to identify issues and concerns and strategies for improvement. The “grade-level” format provides a view of the benchmark difficulty at each grade level, while the “mathematical content progressions” format provides a view of specific content ideas as they develop across the grades. Each format features color-coded benchmarks that correspond to difficulty levels.

Suggested Process

1. **Review the state level mathematics benchmark-item difficulty data information by using both the “grade-level” and “mathematical content progressions” color-coded formats of the standards.** (Documents: *2007 Minnesota K-12 Academic Standards in Mathematics by Grade with Benchmark-item Difficulty Coding* and *2007 Minnesota K-12 Academic Standards in Mathematics by Progressions with Benchmark-item Difficulty Coding*.)

2. **Obtain the necessary School or District Benchmark Reports.** Most schools and districts have a benchmark report that is accessed on the Pearson Access site. The District Assessment Coordinator will be able to provide access to the reports. For more information, see the *2015 Mathematics MCA-III Benchmark Report User Guide* on the Pearson Access site. A picture of the Grade 4 Template is shown on page 4 of the Overview document included in this toolkit. A sample 4th Grade District Benchmark Report is available on page 5. This how-to document provides information on how to use these forms.

3. **Review the *Grade-level Templates*.** The Grade-level templates contain information regarding state-level benchmark difficulty levels for grades 3-8 and high school. Each template shows a list of the benchmarks for a given grade color-coded to the state-level benchmark-item difficulty data. The coding scheme identifies the least difficult benchmarks, the most difficult benchmarks, and the benchmarks that are of average difficulty.

4. **Transfer local school/district data from the *2015 Mathematics MCA-III Benchmark Report* to the table in the *Grade-level template* for each grade.** The *2015 Mathematics MCA-III Benchmark Report* is a tool that can be used to compare the benchmark performance of students in a school relative to their overall performance on the Mathematics MCA-III. That is, a school's performance on each benchmark is described in terms of a deviation around the performance expected given its students' scores on the entire test. (Refer to the *2015 Mathematics MCA-III Benchmark Report Users Guide* for more information. A portion of this user guide is copied at the end of this document for quick access.)

   The columns on the template table provide room to record information to help determine which benchmarks may need further investigation.

   a. The first column is for marking the benchmarks for which student performance was significantly above the expected performance (green squares on the Benchmark Report). Mark the cell for any green squares.
   b. The second column is for marking the benchmarks for which student performance was significantly below the expected performance (red circles on the Benchmark Report). Mark the cell for any red circles.
   c. There is a column to record the highest and lowest performing benchmarks for each strand. Find the benchmark(s) that are the farthest to the right of the expected performance for highest and farthest to the left or the expected performance (this could be the same as a green square or a red circle.).
   d. Finally, there is a column for tracking other information that you may want collect for year to year comparisons. Possible considerations are the benchmarks that have "wide" credible bands (confidence intervals) or two benchmarks that do not overlap. If two credible bands overlap by more than one-half, regardless of color or position of the markers, performance on those benchmarks may be considered statistically equivalent. In other words, if the bands on two different benchmarks have substantial overlap, there is little credible evidence to suggest that actual performance was significantly different on the two benchmarks. If the credible bands across two benchmarks do not overlap, then there is very clear evidence of a reliable difference in performance between the two benchmarks. Mark cells in the template for each grade as appropriate.
   e. Consider adding more columns as needed, including considerations for data that could be tracked over multiple years. Perhaps the triangles that are under expected performance could be watched. Is there a cluster of them from one strand or standard?

5. **Review the state and local data on each Grade-level template and record notes under the "Observations and Conclusions" section.** What observations and conclusions can be drawn from the data? What needs to be explored further? See the questions below to help identify possible issues for improvement. (A sample completed "Grade 4 Template" is provided in the Overview materials on page 5.)

    - How does the state-level benchmark difficulty data compare with the district/school-level data?
    - Does the district/school Benchmark Report data mirror what is happening in classrooms?
    - Are there gaps in the alignment of the curriculum to the standards and benchmarks?
    - If the curriculum is aligned to the standards and benchmarks, how are we ensuring that the curriculum is being taught as intended? (Are we sure that all students are receiving instruction on all benchmarks?)
    - Do students receive instruction in all of the benchmarks before the MCA-III is given?
    - What data are collected at the local level to monitor achievement on the standards and benchmarks?
    - How do teachers record progress toward the standards and benchmarks?
    - What collaborative structures are in place to support adult learning of mathematics content, learning trajectories, and different instructional responses to student thinking?
    - Are beliefs about mathematics or beliefs about learners causing lower expectations, narrower curriculum, and remedial tracks?
    - Is there an overemphasis on learning procedures without connections to meaning, conceptual understanding, and the applications of mathematics that require these procedures?
    - Is there sufficient attention to mathematical problem solving, representations, and reasoning, or do learning experiences and assessments emphasize skills and factual recall?

6. **Investigate other kinds of school- or district-level data.** What other sources of data should be reviewed along with the state and local benchmark difficulty data in order to make decisions about possible improvement actions?

7. **Develop a process to come to consensus on possible improvement actions and develop an improvement plan.** Key strategies might include the following—

    - "Bundle" or cluster the benchmarks across strands to maximize the coherence of mathematics instruction within each grade level. For example, the elementary benchmarks in the *Algebra* strand support the benchmarks in the *Number and Operation* strand and encourage students to develop conceptual connections for procedures. These connections are critical for student achievement in mathematics.
    - Consult the *Minnesota Mathematics and Science Frameworks* (www.scimathmn.org/stemtc) developed by the Minnesota Department of Education and SciMathMN. The Frameworks were written by Minnesota teachers to support

understanding of the Minnesota standards and benchmarks in mathematics and science. Many excellent resources are provided to help teachers translate the standards into classroom practice and assist in student achievement of those standards.

- Use the Achievement Level Descriptors for a check on rigor across the grades. It is very difficult to understand the rigor of the standards at a particular grade level without knowing the standards and benchmarks at least one year below and one year above the grade level. In many cases, when teachers review the Achievement Level Descriptors they recognize the impact of their particular interpretation of the standards. Typical comments are, "What I do in my 4th grade class matches the Achievement Level Descriptors for 3rd grade!" or "How I think about math is keeping my students at 'partially meets'—I don't do much of what is listed in the 'meets or exceeds.'"

A Portion of the *2015 Mathematics MCA-III Benchmark Report Users Guide*

A portion of the *2015 Mathematics MCA-III Benchmark Report Users Guide* is copied below for emphasis and quick reference. Please visit the link above for the complete report.

The primary purpose of the MDE Benchmark Report is to provide information to help curriculum and instructional staff in making inferences about their instructional/curricular activities and their students' level of understanding, based on performance data from the online Mathematics MCA-III. The purpose of data in this report is not to designate strengths and weaknesses in the school. Rather, the Benchmark Report is to serve as a guidance tool to identify possible gaps in instructional content that the school staff may find relevant and important. In particular, it is important to recognize that this report reflects data on a sample of student testing behavior obtained at a single time point in the academic year, and may not fully reflect systemic instructional and curricular outcomes as a whole. Furthermore, some of the results may depend upon the timing and sequence of when content was presented during the school year. For those reasons, it is critical to appropriately involve knowledgeable instructional staff in the discussion and interpretation of the results, and in deliberations about their implications for curriculum and instructional activities.

Evaluating Performance Differences between Benchmarks

In making comparisons between pairs of benchmarks, pay close attention to the amount of overlap of the credible bands for those benchmarks. If their credible bands overlap by more than one-half, regardless of color or position of the markers, performance on those benchmarks may be considered statistically equivalent. In other words, if the bands on two different benchmarks have substantial overlap, there is little credible evidence to suggest that actual performance was significantly different on the two benchmarks. If the credible bands across two benchmarks do not overlap, then there is very clear evidence of a reliable difference in performance between the two benchmarks.

Cautions in Interpreting the Benchmark Report

As with any data, caution must be exercised in making inferences from the benchmark report. It is important to frame any interpretation within the context of the school's environment.

Consideration of external information about the Mathematics curriculum, instructional practices and data from other classroom assessments is critical to making appropriate and meaningful inferences from this report. Interpretation of this report should also take the following factors into account:

- The generalizability of inferences about student performance in the content domain that the benchmark comprises depends upon the representative sampling of: (a) items from the benchmark that students in a school are administered; and (b) students in a school who are administered items from the benchmark. For a computer-adaptive test, such as the online Mathematics MCA-III, there generally will be multiple items administered across students at a school assessing each benchmark.
- Adaptive test blueprint specifications are at the strand and standard level. Thus, benchmark coverage can vary for each student and school. In an adaptive test, this variation in coverage will depend to some degree on ability levels of the students in each school because benchmarks can vary in their inherent difficulty. The length of the credible band around a benchmark report marker reflects, in part, the number of item responses included in calculating the benchmark CLES value; shorter credible bands are associated with larger numbers of student responses to items from the benchmark.
- Different sets of students may be administered items from any particular benchmark. Some benchmarks may have an item by student ability interaction which would result in lower ability students being either over- or under-represented on responses to items from a benchmark.

There are several misinterpretations that should be avoided:

- Color/shape and position of markers in the graphs do not reflect benchmark difficulty.
- Color/shape and position of markers in the graphs do not correspond to achievement levels (i.e., Does Not Meet, Partially Meets, Meets, or Exceeds the Standards).
- When comparing Benchmark Report graphs from different schools within a district, be aware that the range of values on the horizontal axis CLES scale is adjusted to fit each school's data. If a school has a large outlier (i.e., a benchmark with very high or very low relative performance) the graph will have a greater range reflected on the horizontal axis, and its benchmark markers will appear to be clustered more tightly together than those for a school with a smaller range of benchmark CLES values.