

Online Mathematics MCA-III Benchmark Reports

The Online Mathematics MCA-III Benchmark Report is a tool that educators can use to compare the performance of students in their school on benchmarks and standards *relative to their overall performance* on the online Mathematics MCA-III. That is, school performance on each benchmark is described in terms of a deviation around the performance expected based on its students' scores on the entire test. The first section of this document presents an introduction to the benchmark reports and their interpretation. The second section presents a more-detailed discussion of the technical details involved in the calculations that are displayed in the reports.

Overview

The 2014 Grades 3 to 8 reports are based on the online Mathematics MCA-III assessment and organized by the content strands in the Minnesota Academic Standards for Mathematics (2007). A separate graph is produced for each strand to report school performance on the benchmarks that the strand comprises. A list of the benchmarks assessed by the Mathematics MCA-III is included in the document *MCA-III Test Specifications: Mathematics, Grades 3–8(2007 Standards)* available on the MDE website: <http://education.state.mn.us/MDE/EdExc/Testing/TestSpec/>.

Table 1

Mathematics MCA-III Strands

Grades 3 to 5	Grades 6 to 8
Strand 1: Number & Operation Strand 2: Algebra Strand 3: Geometry & Measurement Strand 4: Data Analysis	Strand 1: Number & Operation Strand 2: Algebra Strand 3: Geometry & Measurement Strand 4: Data Analysis & Probability

How to Read the Online Mathematics MCA-III Benchmark Reports

Figure 1 displays the performance for a school on five benchmarks from the grade 6 Algebra strand. Each plotted point represents performance on a benchmark (identified by its numeric label) relative to overall performance expectation for the school (represented by the dotted vertical line that crosses the x-axis at value 0.50). The relative performance on benchmarks for students within a school is reported using the Common Language Effect Size (CLES). When benchmark performance for a group of students is equivalent to that expected given their overall MCA scores, the CLES will equal 0.50. CLES values greater than 0.50 indicate that performance on the benchmark by students at the school exceeds that expected from their overall scores. CLES values less than 0.50 have the opposite implication: school performance on the benchmark is lower than expected based on the overall MCA scores of students in the school.

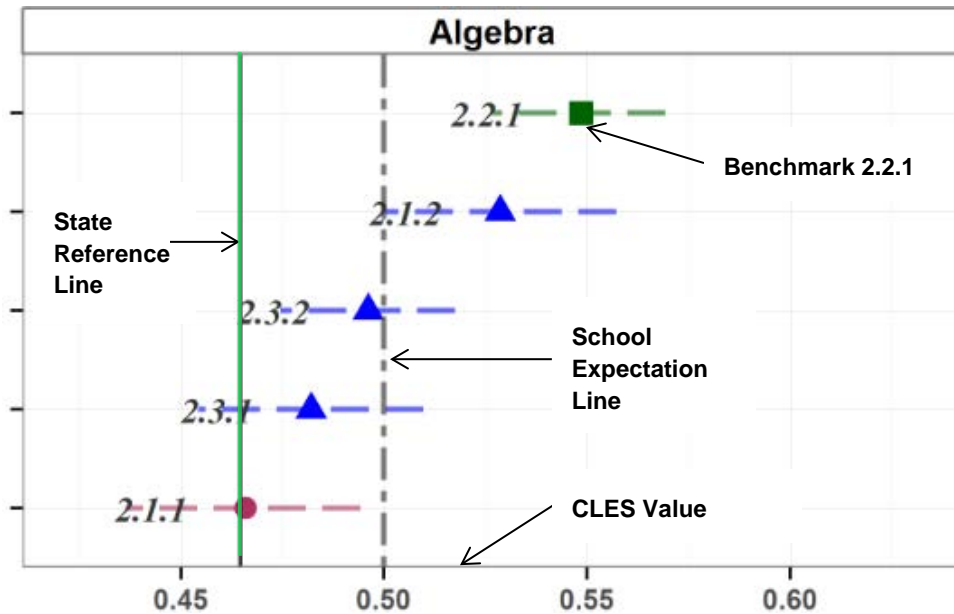


Figure 1. Sample Benchmark Report for the Algebra Strand in Grade 6.

In addition to markers showing the school’s relative performance on each benchmark, a synthetic state-wide reference line (the solid green vertical line) is included in the report to add a normative perspective. This state-wide reference line reflects the performance expected if each student in the school was at the state mean for that grade in overall mathematics ability.

Benchmark Indicators

Individual benchmarks within each strand are identified by a 3-digit code. Relative performance on each benchmark is indicated by a color- and shape-coded symbol and a dotted line extending from the symbol. The symbol’s horizontal position indicates the actual effect size value for the benchmark (on the CLES metric). The dotted line around the symbol represents a corresponding 95% credible interval (i.e., a 0.95 probability range of plausible CLES values given the data). Within each strand, the benchmarks are arranged from highest performance relative to the school at the top right of the graphic to lowest relative performance at the bottom left. As described in Table 2, the color and shape of each plotted symbol indicate how the school’s students performed on the benchmark *relative to expectations* based on the total ability scores of students who were administered items for that benchmark:

Table 2

Benchmark Marker Color and Shape Codes

Green Square: Students performed significantly above expectation on the benchmark. Assigned to markers to the *right* of the dashed vertical line with credible bands that do not overlap this reference line.

Blue Triangle: Students performed near expectation on the benchmark. Assigned to markers with credible bands that overlap the dashed vertical line.

Red Circle: Students performed significantly below expectation on the benchmark. Assigned to markers to the *left* of the dashed vertical line with credible bands that do not overlap this reference line.

The rules described in Table 2 for evaluating benchmark performance relative to school expectation can be used to evaluate school benchmark performance relative to *statewide expected performance* indicated by the solid vertical line:

- Markers to the *right* of the solid vertical line with credible bands that do not overlap this reference line indicate students performed significantly above the expected state-wide performance on the benchmark.
- Markers with credible bands that overlap the solid vertical line indicate students performed near the expected state-wide performance on the benchmark.
- Markers to the *left* of the solid vertical line with credible bands that do not overlap this reference line indicate students performed significantly below the expected state-wide performance on the benchmark.

Evaluating 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 band across two benchmarks does not overlap, then there is very clear evidence of a significant difference in performance between the two benchmarks.

Benchmark Codes

Benchmark codes are indicated to the left of each marker. The benchmark code has been shortened by removing the grade-level indicator. In grade 6, for example, 1.1.4 in the Number & Operation (Strand 1) section corresponds to benchmark 6.1.1.4. In the Strand 4, Data Analysis & Probability section, for example, 4.1.2 corresponds to benchmark 6.4.1.2. A list of the benchmarks assessed by the Mathematics MCA-III is included in the document *MCA-III Test Specifications: Mathematics, Grades 3–8 (2007 Standards)* available on the MDE website: <http://education.state.mn.us/MDE/EdExc/Testing/TestSpec/>.

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.

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 the systematic 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 deliberations about their implications for curriculum and instructional activities.

Technical Details for the Grades 3-8 Online Mathematics MCA-III Benchmark Reports

Relative Benchmark Performance and Common Language Effect Size

The relative performance of benchmarks for students within a given school is reported using the Common Language Effect Size (CLES). The CLES is a non-parametric statistic used to summarize group differences. The basic notion is that two groups (say, Group A and Group B) exist, where each group member has a score on an outcome of interest. The CLES is calculated as the probability that a randomly selected member from one group (e.g., Group A) will have a higher score than a randomly selected member of the other group (Group B). When the groups are equivalent, the probability will be 0.50. As scores in Group A become increasingly higher than those in Group B, the probability that the score of a randomly selected member from Group A will be greater than that of a randomly selected member from Group B increases correspondingly, and the CLES becomes increasingly greater than 0.50. Conversely, as scores in Group A become progressively lower than those in Group B, the CLES will move progressively lower than 0.50.

The Online Mathematics MCA-III Benchmark report uses the CLES to compare performance of two groups on the items administered from a benchmark. The scores comprising the first group are the observed item scores for students in the school on items from the benchmark. The scores comprising the second group are the item scores on benchmark items that would be expected from each student, given their overall score on the Mathematics MCA-III. These expected scores (or *average conditional performance*) are calculated based on the 3-parameter logistic (3PL) measurement model that underlies all scaling on the MCA-III. The 3PL model estimates the probability of a correct response on each benchmark item, given each student's overall MCA-III score.

Here, the School-Based CLES expresses the probability that a student selected at random from the school would receive a higher score on the set of items representing the benchmark than a student drawn at random from a group whose overall MCA-III score distribution was identical to that of the school, but whose benchmark item scores were those expected based on the 3PL model (i.e., average conditional performance). When observed benchmark item response scores are equivalent to those expected based on overall MCA-III scores, the result will be a CLES of 0.50. For each benchmark area, the average conditional performance for a school/district is represented as **0.50** on the School-Based CLES scale, which is interpreted as a student drawn from a random conditional distribution for the school/district having a 50/50 chance of exceeding the expected performance of another student drawn from a random conditional distribution on any particular benchmark. For example, CLES values greater than 0.50 indicate that student performance on the benchmark at the school exceeds the expected conditional student performance. CLES values less than 0.50 have the opposite implication: school performance is lower than expectation given the ability of students that were administered the set of items representing the benchmark.

Within the graph for each strand, a gray vertical dashed line at the 0.50 position on the horizontal axis represents average (expected) school performance for each benchmark area. This expected performance is based on the total ability score of students within the district/school and effectively anchors each benchmark at CLES = 0.50. Thus, the dotted grey vertical dashed line reflects performance on a benchmark right at expectations based on the total test score for all students administered items within any particular benchmark.

State Reference Line

Note that in an adaptive testing format there is no actual state-wide performance as there would be with a fixed form test; no item is administered to all students or a representative sample of all students. The lack of an overall state-wide average is due to students being administered sets of items based on adaptive testing (e.g., selection of items tailored to their ability level). However, an expected state-wide performance *relative to each school* can be estimated and serve as a reference to how well each school did on each benchmark compared to the state average. This state-wide reference line reflects the performance expected if each student in the school was at the state mean for that grade in overall mathematics ability.

95% Credible Interval Bands

The credible interval is the Bayesian analogue of the confidence interval reported in common statistical (i.e., frequentist) practice. The 95% credible interval band reported here can be interpreted as the range of CLES values within which there is a 95% probability the true CLES value lies, given the observed data. The 95% credible interval bands are estimated empirically, based on observed CLES values resulting from 20,000 paired random draws from beta distributions with parameters (1+observed number correct, 1+observed number wrong) for the school observed data and (1+expected number correct, 1+expected number wrong) for the school expectation data. The .025 and .975 quantiles of the observed CLES sampling distribution serve as the limits of the 95% credible interval band. One consequence of this empirical approach is that when a district has a single school at a grade, the sample distributions from school and district analyses can differ very slightly, and the resulting school and district benchmark graphs will be not quite identical.