



## Classroom-Level Data Element Detail

---

Each school's data file includes classroom-level data for analyses for which GOSA flagged the school. The key below explains the information included in each column, details the calculation methodology, and highlights which metrics GOSA uses to generate flags for the Assessment Audit.

### End of Grade Assessments

*Jump to [End of Course Assessments](#)*

#### Answer Change Analysis – End of Grade

Data Recognition Corporation (DRC), the test vendor that administers Georgia Milestones, provides data for wrong-to-right answer changes (erasures for paper-and-pencil tests). The assessment audit analysis excludes classrooms with fewer than eight students.

#### Column Descriptions

|   |            |   |
|---|------------|---|
| A | systemcode | System/Local Education Agency (LEA) code  |
| B | systemname | System/LEA name   |
| C | schoolcode | School code   |
| D | schoolname | School name   |
| E | grade      | Test grade level  |
| F | content    | Test subject area   |
| G | classname  | Test administrator  |
| H | n_students | Number of students in the class<br><i>Data file excludes classes with fewer than eight students from data report.</i> |
| I | n_wtr      | Number of wrong-to-right erasures/changes   |
| J | mean_wtr   | Average number of wrong-to-right erasures/changes for the class   |
| K | std_wtr    | <a href="#">Standard deviation</a> of wrong-to-right erasures/changes for the class                                   |
| L | min_wtr    | Minimum number of wrong-to-right erasures/changes for the class   |
| M | max_wtr    | Maximum number of wrong-to-right erasures/changes for the class   |

N      z\_wtr      Computed standard score of classroom wrong-to-right answer changes:

$$z = \frac{x - \mu}{\sigma/\sqrt{n}}$$

$x$  = mean number of wrong-to-right answer changes for classroom

$\mu$  = mean number of wrong-to-right answer changes of state population

$\sigma$  = standard deviation of state population

$n$  = number of students in classroom

### Answer Change Analysis Flag

For EOGs, GOSA flags schools through the Answer Change Analysis if the school meets at least one of the below criteria:

- Five percent or more of classrooms in a school have a wrong-to-right standard score (Column N) greater than or equal to 4.0 (highlighted in yellow in the *Answer Change – EOG* tab) *OR*
- One classroom in a school has a wrong-to-right standard score (Column N) greater than or equal to 7.0 highlighted in green in the *Answer Change – EOG* tab).

## Unusual Response Pattern Analysis – End of Grade

DRC reports the Unusual Response Pattern analysis for English/Language Arts and Mathematics assessments beginning with 4<sup>th</sup> grade. The calculation is based on the methodology of Jacob and Levitt (2003).<sup>1</sup> The Unusual Response Pattern analysis includes two indices. The first index ranks each classroom's average test score gains relative to other classrooms in that grade and subject, using matched student test data. The ranking is scaled to percentiles (Column L).

The second index includes unexpected patterns in student answers, considered four ways:

- Unlikely blocks of identical answers given by students on consecutive items,
- Degree of correlation in student responses across the test (particularly for unexpected answers),
- Variance of test residuals for the cohort (particular questions having extremely high residual deviations within cohorts), and
- Cases in which students miss easy items while answering difficult items correctly.

DRC then ranks classrooms across each of these four measures. The ranks are combined to create a composite index, scaled to percentile rankings, shown in Column Q.

### Column Descriptions

|   |            |  |
|---|------------|--|
| A | systemcode | System/LEA code  |
| B | systemname | System/LEA name  |
| C | schoolcode | School code  |
| D | schoolname | School name  |
| E | classname  | Test administrator   |
| F | grade      | Test grade level   |
| G | content    | Test subject area  |
| H | form       | Test form  |
| I | n_students | Number of students in the group<br><i>Data file excludes classes with fewer than eight students from data report.</i>  |
| J | tstat      | t-statistic from cohort scale score analysis<br><i>Critical value from significance <a href="#">t-test</a> comparing each classroom administrator's average test score gains to the other administrators' gains for the particular grade and subject</i>   |
| K | pval       | <a href="#">P-value</a> from t-test<br><i>P-value from significance test referenced in Column J.</i> <ul style="list-style-type: none"><li>• 0 = low probability that gains are expected compared to state norms (unusual)</li><li>• 1 = high probability that gains are expected compared to state norms (expected)</li></ul> |

---

<sup>1</sup> Jacob, B. & Levitt, S (2003). Rotten apples: An Investigation of the prevalence and predictors of teacher cheating, *The Quarterly Journal of Economics*, 118 (3), 843-877.

|          |                  |  |
|----------|------------------|--|
| <b>L</b> | <b>index1</b>    | <b>Scale score outlier rank</b><br><i>Uses p-values to rank schools into <a href="#">percentiles</a> so that classrooms with least likely p-values rank closer to 1 and classrooms with expected gains are closer to 0; an index score of 0.95 correlates to the 95<sup>th</sup> percentile</i>                              |
| <b>M</b> | quantregmeasure1 | Quantile regression - Measure 1<br><i>Calculates likelihood of each student's answers based on selected option, number of options, student's previous test scores, and students in the cohort who gave the same response, <a href="#">regressed</a> to account for class size on a scale of 0 (expected) to 1 (unusual)</i>  |
| <b>N</b> | quantregmeasure2 | Quantile regression - Measure 2<br><i>Calculates degree of correlation in student responses across the test, particularly for unexpected/difficult answers, regressed to account for class size on a scale of 0 (expected) to 1 (unusual)</i>  |
| <b>O</b> | quantregmeasure3 | Quantile regression - Measure 3<br><i>Calculates degree of variance of same responses to the same questions, regressed to account for class size on a scale of 0 (expected) to 1 (unusual)</i>   |
| <b>P</b> | quantregmeasure4 | Quantile regression - Measure 4<br><i>Calculates instances in which a student missed easy items while answering difficult items correctly, regressed to account for class size on a scale of 0 (expected) to 1 (unusual)</i>   |
| <b>Q</b> | <b>index2</b>    | <b>Rank of Index 2 scores</b><br><i>Combination of four Measure Ranks (Columns M-P), scaled to percentile rank such that 0 is expected and 1 is unusual; an index score of 0.95 correlates to the 95<sup>th</sup> percentile</i>   |
| <b>R</b> | <b>flag</b>      | <b>Unusual Response Pattern Flag</b><br><i>Indicates if the classroom was in the 95<sup>th</sup> percentile for both indices</i> <ul style="list-style-type: none"> <li>• 0 = Not flagged</li> <li>• 1 = Flagged (Index 1 – Column L <math>\geq</math> 0.95 <u>and</u> Index 2 – Column Q <math>\geq</math> 0.95)</li> </ul> |

### **Unusual Response Pattern Flag**

GOSA flags schools for unusual response patterns if the school meets the below criterion:

- Two or more testing groups in a school were in the 95<sup>th</sup> percentile or greater for unusual responses, meaning the testing group ranked in the 95<sup>th</sup> percentile for both indices (Column L and Column Q; flag shown in Column R). Percentiles for each index of 95 or higher are highlighted in **yellow** in the *Unusual Response Pattern – EOG* tab; the flag is highlighted in **green**.

## Gain Score Analysis – End of Grade

DRC reports the Gain Score Analysis for English/Language Arts and Mathematics assessments beginning with 4<sup>th</sup> grade. Using the classroom administrator of record in 2018, the difference in mean scale scores for the group of students associated with the administrator is compared against the difference in scale scores for the state.<sup>2</sup> Students are matched from 2017 to 2018. Gain score calculations are based on [t-test](#) results between gains in the testing group versus the state.

### Column Descriptions

|   |             |  |
|---|-------------|--|
| A | systemcode  | System/LEA code  |
| B | systemname  | System/LEA name  |
| C | schoolcode  | School code  |
| D | schoolname  | School name  |
| E | grade       | Test grade level   |
| F | content     | Test subject area  |
| G | classname   | Test administrator   |
| H | n_2017      | Number of students in class in 2017  |
| I | mean_2017   | Average scale score in 2017  |
| J | sd_2017     | <a href="#">Standard deviation</a> of scale score in 2017  |
| K | z_mean_2017 | Average standardized score in 2017<br><i>Average scale score (Column I) standardized to account for difference in scale score metrics across the grades</i>  |
| L | n_2018      | Number of students in class in 2018  |
| M | mean_2018   | Average scale score in 2018  |
| N | sd_2018     | Standard deviation of scale score in 2018  |
| O | z_mean_2018 | Average standardized score in 2018<br><i>Average scale score (Column M) standardized to account for difference in scale score metrics across the grades</i>  |
| P | tstat       | t-statistic from cohort scale score analysis<br><i>Critical value from significance t-test comparing the difference between the students in each class's average 2018 score (Column O) and those students' average 2017 scores (Column K) to the difference between the state average in 2018 and 2017</i> |
| Q | p_value     | <a href="#">P-value</a> from significance test referenced in Column P <ul style="list-style-type: none"><li>• 0 = low probability that gains are expected compared to state norms (unusual)</li><li>• 1 = high probability that gains are expected compared to state norms (expected)</li></ul>            |

---

<sup>2</sup> DRC converts the scaled scores to z-scores before computing the statistical test to account for differences in scale scores across grades.

**R**      **outlierscore**

$$OS = \left| 1.0861 \ln \left( \frac{p}{q} \right) \right|$$

$p$  = p-value of the t-test referenced in Column P

$q = 1 - p$

*The coefficient 1.0861 was used such that a p-value of .0001 (significantly low probability that such gains would occur) results in an outlier score of 10. Outlier scores are reported on a scale of 0 to 50. An outlier score of 10 or greater is considered different from the baseline.*

### **Gain Score Analysis Flag**

GOSA flags schools for gain scores if the school meets the below criterion:

- One or more classrooms has an outlier score (Column R) of 10 or greater, highlighted in **yellow** in the *Gain Score – EOG* tab.

## Response Time Analysis – End of Grade

DRC reports the Response Time Analysis for online administrations of the Georgia Milestones. The INSIGHT test administration software captures test duration for all operational test items for each student.

### Column Descriptions

|   |            |   |
|---|------------|---|
| A | systemcode | System/LEA code   |
| B | systemname | System/LEA name   |
| C | schoolcode | School code   |
| D | schoolname | School name   |
| E | grade      | Test grade level  |
| F | content    | Test subject area   |
| G | classname  | Test administrator  |
| H | n_students | Number of students in the class<br><i>Data file excludes classes with fewer than eight students from data report.</i> |
| I | mean       | Average of student test durations in the class, in minutes  |
| J | std        | <a href="#">Standard deviation</a> of student test durations in the class   |
| K | min        | Maximum test duration in the class  |
| L | max        | Minimum test duration in the class  |
| M | zscore     | <b>Computed <a href="#">standard score</a> of classroom test durations:</b>   |

$$= \frac{x - \mu}{\sigma/\sqrt{n}}$$

$x$  = mean test duration for the classroom

$\mu$  = mean test duration for the state population

$\sigma$  = standard deviation of state population

$n$  = number of students in classroom

### Response Time Analysis Flag

For EOGs, GOSA flags schools for response times if the school meets at least one of the below criteria:

- Thirty percent or more of classrooms in a school have a test duration standard score (Column M) greater than or equal to 4.0 or less than or equal to -4.0, highlighted in **yellow** in the *Response Time – EOG* tab.
- One or more classrooms in a school has a test duration standard score (Column M) greater than or equal to 7.0 or less than or equal to -7.0. This criterion identifies classrooms that administered tests for a time period substantially higher or lower than the state norm, highlighted in **green** in the *Response Time – EOG* tab.

# End of Course Assessments

## Answer Change Analysis – End of Course

Data Recognition Corporation (DRC), the test vendor that administers Georgia Milestones, provides data for wrong-to-right answer changes (erasures for paper-and-pencil tests). The assessment audit analysis excludes classrooms with fewer than eight students.

### Column Descriptions

|   |            |   |
|---|------------|---|
| A | systemcode | System/LEA code   |
| B | systemname | System/LEA name   |
| C | schoolcode | School code   |
| D | schoolname | School name   |
| E | content    | Test subject area   |
| F | classname  | Test administrator  |
| G | n_students | Number of students in the class<br><i>Data file excludes classes with fewer than eight students from data report.</i> |
| H | n_wtr      | Number of wrong-to-right erasures/changes   |
| I | mean_wtr   | Average number of wrong-to-right erasures/changes for the class   |
| J | std_wtr    | <a href="#">Standard deviation</a> of wrong-to-right erasures/changes for the class                                   |
| K | min_wtr    | Minimum number of wrong-to-right erasures/changes for the class   |
| L | max_wtr    | Maximum number of wrong-to-right erasures/changes for the class   |
| M | z_wtr      | Computed <a href="#">standard score</a> of classroom wrong-to-right answer changes:                                   |

$$z = \frac{x - \mu}{\sigma/\sqrt{n}}$$

$x$  = mean number of wrong-to-right answer changes for classroom

$\mu$  = mean number of wrong-to-right answer changes of state population

$\sigma$  = standard deviation of state population

$n$  = number of students in classroom

### Answer Change Analysis Flag

For EOCs, GOSA flags schools through the Answer Change Analysis if the school meets at least one of the below criteria:

- Five percent or more of classrooms in a school have a wrong-to-right standard score (Column M) greater than or equal to 5.0 (highlighted in **yellow** in the *Answer Change – EOC* tab) *OR*
- One classroom in a school has a wrong-to-right standard score (Column M) greater than or equal to 7.0 highlighted in **green** in the *Answer Change – EOC* tab).



## Unusual Response Pattern Analysis – End of Course

DRC reports the Unusual Response Pattern analysis for English/Language Arts and Mathematics assessments beginning with 4<sup>th</sup> grade. The calculation is based on the methodology of Jacob and Levitt (2003).<sup>3</sup> The Unusual Response Pattern analysis includes two indices. The first index ranks each classrooms' average test score gains relative to other classrooms in that grade and subject, using matched student test data. The ranking is scaled to percentiles (Column K).

The second index includes unexpected patterns in student answers, considered four ways:

- Unlikely blocks of identical answers given by students on consecutive items,
- Degree of correlation in student responses across the test (particularly for unexpected answers),
- Variance of test residuals for the cohort (particular questions having extremely high residual deviations within cohorts), and
- Cases in which students miss easy items while answering difficult items correctly.

DRC then ranks classrooms across each of these four measures. The ranks are combined to create a composite index, scaled to percentile rankings, shown in Column P.

### Column Descriptions

|   |            |  |
|---|------------|--|
| A | systemcode | System/LEA code  |
| B | systemname | System/LEA name  |
| C | schoolcode | School code  |
| D | schoolname | School name  |
| E | classname  | Test administrator   |
| F | content    | Test subject area  |
| G | form       | Test form  |
| H | n_students | Number of students in the group<br><i>Data file excludes classes with fewer than eight students from data report.</i>  |
| I | tstat      | t-statistic from cohort scale score analysis<br><i>Critical value from significance <a href="#">t-test</a> comparing each classroom administrator's average test score gains to the other administrators' gains for the particular grade and subject</i>   |
| J | pval       | <a href="#">P-value</a> from t-test<br><i>P-value from significance test referenced in Column I.</i> <ul style="list-style-type: none"><li>• 0 = low probability that gains are expected compared to state norms (unusual)</li><li>• 1 = high probability that gains are expected compared to state norms (expected)</li></ul> |

---

<sup>3</sup> Jacob, B. & Levitt, S (2003). Rotten apples: An Investigation of the prevalence and predictors of teacher cheating, *The Quarterly Journal of Economics*, 118 (3), 843-877.

|          |                  |  |
|----------|------------------|--|
| <b>K</b> | <b>index1</b>    | <b>Scale score outlier rank</b><br><i>Uses p-values to rank schools into <a href="#">percentiles</a> so that classrooms with least likely p-values rank closer to 1 and classrooms with expected gains are closer to 0; an index score of 0.95 correlates to the 95<sup>th</sup> percentile</i>                              |
| L        | quantregmeasure1 | Quantile regression - Measure 1<br><i>Calculates likelihood of each student's answers based on selected option, number of options, student's previous test scores, and students in the cohort who gave the same response, <a href="#">regressed</a> to account for class size on a scale of 0 (expected) to 1 (unusual)</i>  |
| M        | quantregmeasure2 | Quantile regression - Measure 2<br><i>Calculates degree of correlation in student responses across the test, particularly for unexpected/difficult answers, regressed to account for class size on a scale of 0 (expected) to 1 (unusual)</i>  |
| N        | quantregmeasure3 | Quantile regression - Measure 3<br><i>Calculates degree of variance of same responses to the same questions, regressed to account for class size on a scale of 0 (expected) to 1 (unusual)</i>   |
| O        | quantregmeasure4 | Quantile regression - Measure 4<br><i>Calculates instances in which a student missed easy items while answering difficult items correctly, regressed to account for class size on a scale of 0 (expected) to 1 (unusual)</i>   |
| <b>P</b> | <b>index2</b>    | <b>Rank of Index 2 scores</b><br><i>Combination of four Measure Ranks (Columns L-O), scaled to percentile rank such that 0 is expected and 1 is unusual; an index score of 0.95 correlates to the 95<sup>th</sup> percentile</i>   |
| <b>Q</b> | <b>flag</b>      | <b>Unusual Response Pattern Flag</b><br><i>Indicates if the classroom was in the 95<sup>th</sup> percentile for both indices</i> <ul style="list-style-type: none"> <li>• 0 = Not flagged</li> <li>• 1 = Flagged (Index 1 – Column L <math>\geq</math> 0.95 <u>and</u> Index 2 – Column P <math>\geq</math> 0.95)</li> </ul> |

### **Unusual Response Pattern Flag**

GOSA flags schools for unusual response patterns if the school meets the below criterion:

- Two or more testing groups in a school were in the 95<sup>th</sup> percentile or greater for unusual responses, meaning the testing group ranked in the 95<sup>th</sup> percentile for both indices (Column K and Column P; flag shown in Column Q). Percentiles for each index of 95 or higher are highlighted in **yellow** in the *Unusual Response Pattern – EOG* tab; the flag is highlighted in **green**.

## Gain Score Analysis – End of Course

DRC reports the Gain Score Analysis for English/Language Arts and Mathematics assessments beginning with 4<sup>th</sup> grade. Using the classroom administrator of record in 2018, the difference in mean scale scores for the group of students associated with the administrator is compared against the difference in scale scores for the state.<sup>4</sup> Students are matched from 2017 to 2018. Gain score calculations are based on [t-test](#) results between gains in the testing group versus the state.

### Column Descriptions

|   |             |  |
|---|-------------|--|
| A | systemcode  | System/LEA code  |
| B | systemname  | System/LEA name  |
| C | schoolcode  | School code  |
| D | schoolname  | School name  |
| E | content     | Test subject area  |
| F | classname   | Test administrator   |
| G | n_2017      | Number of students in class in 2017  |
| H | mean_2017   | Average scale score in 2017  |
| I | sd_2017     | <a href="#">Standard deviation</a> of scale score in 2017  |
| J | z_mean_2017 | Average standardized score in 2017<br><i>Average scale score (Column H) standardized to account for difference in scale score metrics across the grades</i>  |
| K | n_2018      | Number of students in class in 2018  |
| L | mean_2018   | Average scale score in 2018  |
| M | sd_2018     | Standard deviation of scale score in 2018  |
| N | z_mean_2018 | Average standardized score in 2018<br><i>Average scale score (Column L) standardized to account for difference in scale score metrics across the grades</i>  |
| O | t-stat      | t-statistic from cohort scale score analysis<br><i>Critical value from significance t-test comparing the difference between the students in each class's average 2018 score (Column N) and those students' average 2017 scores (Column J) to the difference between the state average in 2018 and 2017</i> |
| P | p_value     | <a href="#">P-value</a> from significance test referenced in Column O <ul style="list-style-type: none"><li>• 0 = low probability that gains are expected compared to state norms (unusual)</li><li>• 1 = high probability that gains are expected compared to state norms (expected)</li></ul>            |

---

<sup>4</sup> DRC converts the scaled scores to z-scores before computing the statistical test to account for differences in scale scores across grades.

**Q**      **outlierscore**

$$OS = \left| 1.0861 \ln \left( \frac{p}{q} \right) \right|$$

$p$  = p-value of the t-test referenced in Column O

$$q = 1 - p$$

*The coefficient 1.0861 was used such that a p-value of .0001 (significantly low probability that such gains would occur) results in an outlier score of 10. Outlier scores are reported on a scale of 0 to 50. An outlier score of 10 or greater is considered different from the baseline.*

### **Gain Score Analysis Flag**

GOSA flags schools for gain scores if the school meets the below criterion:

- One or more classrooms has an outlier score (Column Q) of 10 or greater, highlighted in **yellow** in the *Gain Score – EOC* tab.

## Response Time Analysis – End of Course

DRC reports the Response Time Analysis for online administrations of the Georgia Milestones. The INSIGHT test administration software captures test duration for all operational test items for each student.

### Column Descriptions

|   |            |   |
|---|------------|---|
| A | systemcode | System/LEA code   |
| B | systemname | System/LEA name   |
| C | schoolcode | School code   |
| D | schoolname | School name   |
| E | content    | Test subject area   |
| F | classname  | Test administrator  |
| G | n_students | Number of students in the class<br><i>Data file excludes classes with fewer than eight students from data report.</i> |
| H | mean       | Average of student test durations in the class, in minutes  |
| I | std        | <a href="#">Standard deviation</a> of student test durations in the class   |
| J | min        | Maximum test duration in the class  |
| K | max        | Minimum test duration in the class  |
| L | zscore     | <b>Computed <a href="#">standard score</a> of classroom test durations:</b>   |

$$= \frac{x - \mu}{\sigma/\sqrt{n}}$$

$x$  = mean test duration for the classroom

$\mu$  = mean test duration for the state population

$\sigma$  = standard deviation of state population

$n$  = number of students in classroom

### Response Time Analysis Flag

For EOCs, GOSA flags schools for response times if the school meets at least one of the below criteria:

- Thirty percent or more of classrooms in a school have a test duration standard score (Column L) greater than or equal to 5.0 or less than or equal to -5.0, highlighted in **yellow** in the *Response Time – EOC* tab.
- One or more classrooms in a school has a test duration standard score (Column L) greater than or equal to 7.0 or less than or equal to -7.0. This criterion identifies classrooms that administered tests for a time period substantially higher or lower than the state norm, highlighted in **green** in the *Response Time – EOC* tab.