



Georgia Milestones Assessment System

Mode Comparability Analyses

Executive Summary

The Georgia Milestones Assessment System (Georgia Milestones) is a comprehensive summative system of assessments that span grades 3 through high school. Georgia Milestones measures how well students have learned the knowledge and skills outlined in the state-adopted content standards in English language arts, mathematics, science, and social studies. Students in grades 3 through 8 take an End-of-Grade (EOG) assessment in each content area, while high school students take an End-of-Course (EOC) assessment for each of the ten courses designated by the State Board of Education. Features of Georgia Milestones include:

- open-ended (constructed-response) items in English language arts and mathematics (all grades and courses);
- a writing component (in response to passages read by students) at every grade level and course within the English language arts assessment;
- norm-referenced items in all content areas and courses, to complement the criterion-referenced information and to provide a national comparison; and,
- transition to online administration over time, with online administration considered the primary mode of administration and paper-pencil as back-up until the transition is complete.

The transition to online testing is an important component of the Georgia Milestones program in that it provides the opportunity for students to take their assessments using technology increasingly found at home and in the classroom. Students have consistently indicated that they prefer to take their tests online as opposed to the traditional practice of paper and pencil, as demonstrated through an online survey of students conducted by the Georgia Department of Education in Spring 2015.

As part of the launch of the Georgia Milestones assessments, a series of mode comparability analyses (i.e., examining the impact of the mode of administration – online or paper – on the performance of the test items and students tested) were implemented in two phases. These analyses were conducted in partnership with our Technical Advisory Committee (TAC), a group comprised of six internationally renowned measurement experts. TAC is charged with providing the Georgia Department of Education with impartial review and advice regarding all aspects of Georgia’s student assessment program.

Before describing the analyses conducted after the first administration of the Georgia Milestones EOG and EOC measures, it is important to note that Georgia has a long history of online testing, beginning in 2003. The End of Course Tests (EOCT), launched in 2003, included an online option. Additionally, the Criterion-Referenced Competency Tests (CRCT), launched in 2000, offered the retest online beginning in 2009. Neither of these two testing programs exhibited mode differences.

Nonetheless, with the implementation of Georgia Milestones, it was important to investigate whether this new assessment program exhibited mode differences. The first phase of mode comparability analyses for Georgia Milestones centered on comparing test performance of carefully matched groups of students who took the tests online and via paper. Groups of students were matched on the basis of previous achievement data as well as student background characteristics such as ethnicity and gender. Statistical tests were conducted to evaluate the magnitude of the differences in test scores between the two groups of students.

The results from the mode comparability analyses of the Winter 2014 EOC data found negligible differences between matched groups of students who took the test via different modes. Small, non-negligible differences in test performance by mode were found in some grades/subjects/courses using data from the Spring 2015 EOG and EOC administrations. It is important to note that the differences in test performance by mode were found in both directions, meaning that there was no finding that students were consistently advantaged or disadvantaged by the mode of administration (online or paper) across the tests that comprise the Georgia Milestones Assessment System.

While some *small*, non-negligible differences were found in the Spring, this was compounded by the fact that a substantial proportion of students with disabilities participated in online testing. Twice as many students with disabilities took the tests online than are present in the full population of Georgia students.

Because of these small, non-negligible findings, the second phase of the mode comparability analyses centered on evaluating the robustness of the statistical models used to score and report the Georgia Milestones results with respect to testing mode. This involved a series of

analyses to evaluate mode differences established in Phase 1 by item type/characteristic as well as the model and data fit with respect to test mode. The results from the Phase 2 analyses found that the few differences detected between modes could not be attributed to specific item types (e.g., constructed response) or specific item characteristics (e.g., items that included graphics or other features). An extensive evaluation of the fit to the Georgia Milestones data indicated that the statistical models used to derive student scores were well estimated and functioning appropriately.

As Georgia's Technical Advisory Committee considered the findings across the analyses in both Phases 1 and 2, they concluded that adjustments to scores for mode differences were not warranted. Ongoing evaluation of score comparability will be pursued on a regular basis within Georgia Milestones as the program fully transitions to online testing.