



# *Research Brief*

## *Assessment, Research, and Data Analysis*

---

Vol. 1301  
September 2013

Ms. Felicia Mallory  
Dr. Terry Froman  
Dr. Aleksandr Shneyderman

### **Interim Assessment: Development, Item Analysis, and Prediction**

#### **Background**

The initial decision to develop the Interim Assessment dates back to the 2005-2006 school year, when the School Board contracted with Educational Testing Services (ETS) for the development and delivery of Interim Assessments in Reading, Mathematics, and Science, an item bank, a series of professional development workshops for teachers, and a predictive modeling study. The item bank was rolled out in the 2007-2008 school year with over 20,000 items in Reading and Mathematics for grades 3-10. Science items were added in 2008-2009. In the summer of 2007 (2008 for Science) ETS conducted an item analysis and a predictive study to determine how well the Interim Assessment test items were functioning and to determine the relationship between students' achievement on the Interim Assessment Test and Florida Comprehensive Assessment Test (FCAT). Results of the item analyses indicated that most items functioned well, while some items had problems; these items were removed or replaced. An empirical logistic regression using only raw scores of the common items between 2006-2007 and 2007-2008 was used to predict the achievement levels 3-5 vs. 1-2 on the 2007 FCAT (2008 for Science). ETS used the results from the prediction study to classify students into three categories of satisfactory, limited, and insufficient progress. Students scoring in the satisfactory range were most likely to be on-track for scoring within achievement levels 3-5 in the FCAT, while those in the insufficient range were not likely to be on-track to score that high.

In 2009-2010, the Florida Department of Education released the Next Generation Sunshine State Standards (NGSSS). Miami-Dade County Public Schools, in an addendum to the contract, agreed to have ETS conduct a realignment study of the test items to determine their adherence to the new standards. Although some of the items met the Next Generation Sunshine State Standards, new items in Reading and Mathematics had to be developed. Additionally, items for Interim Assessment corresponding to the state-mandated end-of-course (EOC) assessments had to be developed. Due to monetary cut-backs the district's Student Assessment and Education Testing and the Curriculum and Instruction offices assumed the tasks of project management, production, and test development that were initially undertaken by ETS. Beginning in 2009-2010, new test forms that covered the Next Generation Sunshine State Standards were created from the secure item bank in Reading and Mathematics. In Science, M-DCPS contracted with Core K-12, a test development company, to produce Interim Assessment

test forms for grades 5 and 8, as well as Biology. M-DCPS Department of Social Sciences managed the work of developing items for Interim Assessments in United States History and Civics.

Currently, the Miami-Dade County Public Schools Mathematics, Reading, Science, and Social Science Interim Assessments are formative benchmark assessments that are based on the M-DCPS instructional pacing guides, aligned to the Next Generation Sunshine State Standards, and adhere to the Florida Department of Education Item Specifications. The Interim Assessments cover Reading in Grades 3-10, Mathematics in Grades 3-8, Algebra I, Geometry, Science for Grades 5, 8, Biology, United States History, and Civics. The assessments are administered three times a year: baseline at the beginning of a school year, in fall, and winter.

Since the development of the new Interim Assessment tests aligned with the Next Generation Sunshine State Standards, the Department of Research Services has conducted an item analysis and the prediction study resulting in a creation of three performance categories for each of the tests. These two tasks are described in the sections below.

## Rasch Item Analysis

Analyzing test data according to the Rasch model, a special case of item response theory, gives a range of details concerning test characteristics and performance. For every interim test in each administration, content area, and grade level, a Rasch analysis was performed. Of particular interest were the following categories of results.

1. **An Item/Person Map.** Matched graphs of the item difficulties and the person abilities allow for assessing the effective operating range and shape of these distributions as well as the alignment of difficulty to ability throughout the test span.
2. **Summaries of Person and Item Statistics.** These tables present general information regarding the measures of central tendency and variation and reliability of the test.
3. **Item Fit Statistics.** Each test item is evaluated for its correlation to the overall score as well as evidence of unusual response patterns which may identify faulty item construction.
4. **Distractor Analysis.** This examination presents confirmation that the choices for the individual items are being responded to in a meaningful manner and possesses convergent validity.
5. **Differential item Functioning.** Each item is evaluated to determine if it shows indications of preferential response patterns with respect to student free/reduced price lunch status, gender, race, and English language proficiency.

After reviewing the results of the Rasch analysis for each test, items flagged as suspicious (either poorly fitting or potentially exhibiting differential item functioning) were re-evaluated and either re-written or eliminated from the final test form. Each person's total test score was then recalculated based on the refined test form and the results were then prepared for subsequent analysis.

## Prediction Study

For each of the fall and spring forms of the Reading, Mathematics, Science, and Biology Interim Assessment tests, a prediction study has been completed. An item analysis of the Interim Assessments in Social Sciences will be conducted during the 2013-2014 school year, and a prediction study for these tests will be conducted following the release of the 2014 US History EOC exam.

For each of the forms of the Interim Assessment test, the prediction study was conducted in several steps. First, a correlation between the raw score on an Interim Assessment and a scale score on the corresponding component of the FCAT or EOC assessment was found. All such correlations were found to be sufficiently high (generally, around .8) to warrant further analyses. Then, the achievement levels on the FCAT or EOC assessment were dichotomized with 0 indicating achievement levels 1-2, and 1 indicating achievement levels 3-5. The binary logistic regression with the dichotomous achievement level as the outcome and the raw score on the interim test as a single predictor was used to calculate the probability of scoring within achievement levels 3-5 for each grade level separately. In every case, the coefficient for the interim test raw score in the regression equation was found to be significantly different from zero ( $p < .01$ ). After the predicted probabilities were calculated, the (0, 1) range was recoded into the three categories based on the following rule:

$p \leq .4$	<i>insufficient progress</i>
$0.4 < p < 0.6$	<i>limited progress</i>
$p \geq 0.6$	<i>satisfactory progress.</i>

Subsequent studies revealed that the prediction was fairly accurate. As an example, the percentage of incorrect predictions based on one of the forms of the Mathematics Interim Assessment varied from 13% to 16%. In Reading, these percentages varied from 10% to 16%.

## Summary

The Miami-Dade County Public School System administers Interim Assessments in Reading, Mathematics, Algebra I, Geometry, Science, Biology, United States History, and Civics. As part of the development process, the Department of Research Services conducts item analyses and prediction studies for each new test form. The Rasch item analysis supplies a range of statistics that help identify items needing special attention. The prediction analysis relates the interim test performance to probable success on the Florida Comprehensive Assessment Test or End-of-Course Assessment in the corresponding content areas. The successful implementation of these analyses ensures that the interim tests exhibit appropriate test characteristics and afford sufficient predictive validity to provide students, teachers, and administrators with useful feedback and pedagogical guidance.