

External Review Questions

Question		Questions Posed by District Representatives on July 29, 2010
Are students accurately matched to their testbook/results?	<ul style="list-style-type: none"> • Student Document Verification Process • Random sample of check between test results records and physical test books 	Group did not believe that this was an issue but recommended actions to help those who have concerns. Some score inquiry requests are being held for now; expect many more.
Are more students showing dramatic changes in test results/achievement level changes?	<ul style="list-style-type: none"> • Comparison percent of students with 1 and 2 or more shift in Achievement Levels (see Analysis of Movement) 	
1. Do the test construction specifications clearly define the design and statistical requirements that would ensure parallel versions of the FCAT assessments across years?	<ul style="list-style-type: none"> • FCAT 2010 Test Construction Specifications/Checklists <p>Sample Provided: Test Construction Checklists for Gr. 4 and Gr. 5 Reading</p>	
2. Is there sufficient evidence that the specifications for developing parallel versions of the FCAT Reading and Mathematics assessments yielded the intended outcomes from 2007 through 2010?	<ul style="list-style-type: none"> • 2007-2010 FCAT Content Representation • FCAT Mathematics Test Characteristic Curves (TCCs) 2007 through 2010 • FCAT Reading TCCs 2007 through 2010 • Conditional standard error of measurement at the scale score Achievement Level cut scores for reading and math across all grades 2007 through 2010. • Conditional standard error of measurement at the developmental scale score Achievement Level cut scores for reading and math across all grades 2007 through 2010. • Standard Error of Measurement curves with Achievement Level cut scores. 	<ul style="list-style-type: none"> • <i>Collect and compare benchmarks/ content focus for grades 4 and 5 reading 2007 – 2010.</i> • <i>Are test construction curves based on field test results?</i> • <i>Are there fewer items at low end of scale?</i> • <i>Around cut points, were the discrimination parameters the same this year as in past years?</i>

	Sample Provided: Gr. 4 and Gr. 5 FCAT Reading and Mathematics Content Representation	
3. Do the scaling and equating specifications clearly define the requirements for producing equivalent score scales across years, including standards for the use of anchor items?	<ul style="list-style-type: none"> 2007-2010 Scaling & Equating Specifications <p>Sample Provided: Scaling and Equating Roles and Responsibilities</p>	.
4. Is there sufficient evidence that the specifications were followed and yielded equivalent score scales from 2007 through 2010?	<ul style="list-style-type: none"> 2007-2010 Scaling & Equating Summary Sheets FCAT Mathematics Post Equating Curves 07-10 FCAT Mathematics Post Equating Curves 07-10 Scatter-plots summarizing the relationship between IRT old and new item parameters by passage, as well as the standard deviation of changes in difficulty (B parameter). Pre-equated TCCs with Achievement Level cut scores. Post-equated TCCs with Achievement Level cut scores. Comparison of student performance in the calibration sample vs. population on the core items across all grades in reading and mathematics 2007 through 2010. 	<ul style="list-style-type: none"> <i>Do we have equating error estimates over time? Is equating error this year different than it has been in the past?</i> <i>Is there an empirical test to make sure the scales are not operating differently this year?</i> <i>What are implications of differences between pre- and post-equating curves? Please compare pre- and post-equating TCCs and describe differences and implications, including relationship to achievement level cutscores. Can these differences impact gain scores?.</i> <i>Are the Achievement Levels set on the calibration sample or does the Department wait until all the scores are in?</i> <i>Please compare calibration sample to state characteristics 2007 – 2010.</i>
5. Are the scale score and learning gains fluctuations from 2008 to 2009 and 2009 to 2010 within or beyond those that should be expected?	<ul style="list-style-type: none"> FCAT Mathematics Trends in Scale Scores FCAT Reading Trends in Scale Scores Percent of schools that lost points on each of the school grades learning gains components over time, by magnitude of the increase or decline. Percentage of Elementary, Middle, and High Schools that lost points on the Reading and 	<ul style="list-style-type: none"> <i>Have we looked at large drops? Please provide data on changes in learning gains across years for large drops (more than 10 points) rather than any drop.</i> <i>Are gains measures a function of schools or margin of error? Is the growth measure working how we anticipated? Are we seeing</i>

	<p>Mathematics Learning Gains Component of School Grades 2002-2010</p> <ul style="list-style-type: none"> • Percentage of Elementary, Middle, and High Schools that lost points on the Reading and Mathematics Learning Gains of the Low 25% component of School Grades 2002-2010 • Average Developmental Scale Score (DSS) Change for All Students and Students in Low 25% by Achievement Level and Grade – Reading and Mathematics • Post-equated TCCs with Achievement Level cut scores. The ability to discriminate at the cut points can be monitored using these curves. • Equipercentile analysis [NOTE: The data and the code were received, several phone calls made between FDOE and Palm Beach School District to clarify the follow up questions. Currently, there are some technical difficulties, such as the version of SAS and error messages. We expect that these issues will be resolved by the end of today and the analyses will be completed by Friday.] • Regression Analyses – See HumRRO report 	<p><i>differences because the number of items at the low end is always so limited?</i></p> <ul style="list-style-type: none"> • <i>If gains are just an unstable measure, should we be using it as primary component of accountability system? Is it fair to hold schools accountable? What could cause the instability in the system?</i> • <i>Please follow up and further analyze the school grades components regarding achievement level changes and DSS changes. This can be done with a regression analysis.</i> • <i>Please replicate the equipercentile analysis conducted by Palm Beach using the state sample.</i>
<p>6. Has the composition of Florida students changed in such a way as to impact student learning gains calculations?</p>	<ul style="list-style-type: none"> • Percent of Free/Reduced Lunch (FRL), English Language Learners and Students with Disabilities among FCAT test takers • FRL Analysis of Gains by Cohorts of Students 	<ul style="list-style-type: none"> • <i>Are students who previously were NOT on FRL performing at a lower level?</i>

<p>7. Has the composition of Florida students in lowest quarter changed in such a way as to impact student learning gains calculations?</p>	<ul style="list-style-type: none"> • Composition of the Low 25% in Reading for Elementary Schools 2007 vs 2008, 2008 vs. 2009, and 2009 vs. 2010 • Average Baseline Developmental Scale Score (DSS) for Level 1 Students in the Low 25%, by Grade 2007 vs. 2008, 2008 vs. 2009, 2009 vs 2010 	<ul style="list-style-type: none"> • <i>Please conduct an analysis examining comparable students with equivalent pre-test scores – rather than just lowest 25% whose composition can change over time.</i>
<p>8. Has the relationship between previous and current developmental scale scores changed over time? If so, are there anomalies? If so, what are these and what is the cause(s)?</p>	<ul style="list-style-type: none"> • Regression Analyses – See HumRRO report 	
<p>9. In 2010, new item types were field tested in all except anchor test forms for grades 3 – 8 Mathematics. Is there evidence that the presence of these new item types had an effect on student achievement?</p>	<ul style="list-style-type: none"> • FCAT 2010 Grade 5 reading data: Analyses of scale score means and standard deviations across anchor and field-test forms. • Significance tests and effect sizes for scale scores between students with field test forms vs students with anchor forms for those grades with state-wide spiraling of forms. • For FCAT 2009 and 2010 reading across all grades, “passage specific pre- and post-equated curves” were drawn. The average position shift of each passage and the year that anchor passages are coming from are noted on the graphs. • Summary of anchor item characteristics. 	<p><i>Did the calibration forms have any field test items?</i></p> <p><i>Was there anything different in operation this year than past years? Was there a significant difference in 4th grade reading last year between forms, given there was no population change.</i></p> <p><i>What could be causing substantial changes in gains when proficiency was at or above last year?</i></p>