

**APPENDIX A**  
**Planning Meeting Agenda and Minutes**

**October 15, 2008**

## Early Elementary Mathematics Assessment Planning Meeting

### AGENDA

FDOE Room 503 ♦ October 15, 2008 ♦ 3:00 – 5:00 PM

FSU		FDOE	
Karen DeMeester ✓	Kris Ellington ✓	Victoria Ash ✗	Denise Bishop ✓
Laura Lang ✓	Chris Johnson ✓	Todd Clark ✓	Shan Goff ✓
Faye Jones ✓	Chris Lonigan ✓	Tara Huls ✓	Cornelia Orr ✗
Beth Phillips ✓	David Purpura ✓	Mary Jane Tappen ✓	Mario Zuniga ✓
Rob Schoen ✗		Tamika Brinson ✓	
		<b>UCF Project CENTRAL (via phone)</b>	
		Mary Little ✓	Shelby Robertson ✓

Topic	Intent	Notes
2008-09 Project Plan	Review to ensure common understanding	
Questions	Discuss, answer, and/or assign to first webinar with Advisory Board	
Advisory Board Membership	Finalize list for invitation	
Literature Review	Discuss focus	
Next Steps	“Touch points” for FDOE?	

Questions	For Advisory Board
What will be the primary use of results? To inform instruction	
Aggregate in PMRN (for what purpose?)	
Other uses?	
IPDP and professional development	
Evaluate standards-based instruction efforts	
Evaluate mathematics program effectiveness	
Monitoring outcomes of change efforts	
Predict FCAT Grade 3 proficiency	
Screen for ESE	
Who will be assessed (All students including SWD, ELL; representative sample)?	
Special versions (e.g., uncontracted Braille, large print)?	

Intended to be “true” formative assessment vs. progress monitoring probes vs. interim assessments? Clarify definitions for common understanding.	
Limited number of “probes” (robust indicators) vs. content sampled/comprehensive?	
Access to new “FCAT” test item specifications for grades 3 – 5?	
Were learning progressions built into the design of NGSSS?	
Computer vs. Hand-Held vs. Paper Format	
Who will administer the assessments?	
What are possible item formats? Checklists, manipulatives, constructed response, ...	
Has content to be measured been determined?	
Number Sense (2?)	
Geometry	
Measurement	
Number of administration periods per school year? (daily → weekly → ...)	
Will there be flexibility in the order of administration for content that is not on a learning progression (order numbers vs. describe geometric shapes)	

### Meeting Minutes

**Purpose:** To discuss expectations and key issues related to FCR-STEM’s development of a plan for a PreK-3 mathematics formative assessment for Florida.

**Purpose of Assessment:** To inform instruction, i.e., help teachers (1) identify what students know and can do and (2) decide on next steps for instruction. Initial conversations with FDOE focused on a formative assessment that would be quickly administered and scored and provide information to teachers on what to do next.

**FCRR’s VPK math assessment:** FCRR will be developing a math formative assessment for the Voluntary PreK (VPK) program. It will be a broad assessment for use by VPK teachers. David Purpura will have a key role in this project which will be the focus of his dissertation. An implementation study will be conducted next year (2009-2010). The VPK assessment is one year ahead of the schedule for the PreK-3 assessment to be developed by FCR-STEM.

**Design and programming:**

**Web-based:** Chancellor Haithcock expects the math formative assessment to be web-based. Palms are too restrictive and the technology is quickly outdated.

**Aggregation/reporting:** Results will be aggregated into an electronic reporting network for teachers (the Reading PMRN or other network). There are advantages to having a separate network for math:

- The Reading PMRN was just redesigned and reprogrammed. It is not timely to add math at this point.
- Reading First drove the design of the PMRN. A separate network for math would offer more flexibility. Although the VPK network will be aligned with the PMRN, it will have a different server/website and a simpler look and feel (e.g., no “box and whiskers” plots).

Alignment: Alignment with the new standards database might be a better option than alignment with the PMRN developed by Infinity. The standards database, the course code directory and curriculum planning tool for math and science have been developed by a different programmer and at a lower cost (although there may be reasons for the higher cost of the PMRN, such as firewalls to protect student confidentiality).

Mary Jane's primary concerns are:

- ease-of-use by teachers not who builds the system or where it is housed. Two different groups building these systems (especially if they're competitors) could be a problem. Teachers may need to learn two different systems.
- alignment of content for the progression of skills across the grade levels. Isolating grades in assessments won't work since students may be below, or progress beyond, the enrolled grade level.

A decision on design, programming, and cost will need to be made soon for a legislative budget request.

**Use of Results:** Assessment results are expected to be used:

1. To inform instruction in the classroom, i.e., help teachers (1) identify what students know and can do and (2) decide on next steps for instruction.
2. To develop IPDPs and identify teacher professional development needs.
3. To evaluate standards-based instruction (to determine whether content standards are being taught to the level that they should be taught).
4. To monitor outcomes of change or reform efforts.
5. To predict FCAT Grade 3 proficiency/probability of success (This is important to the Commissioner)
6. To measure the effectiveness of RTI/continuous progress monitoring
7. To screen for ESE and students who can benefit from early acceleration.
8. To evaluate mathematics program effectiveness (however, the assessment will not be developed to evaluate any specific curriculum or program).
9. As part of the VPK accountability system (e.g., as a measure of student readiness for kindergarten)
10. As part of the performance pay system for K-2 teachers.

*All of the above uses should be discussed with the Advisory Board.*

Issues and questions:

- Some of the above uses will raise standards for psychometric quality which, in turn, will increase the timeline for development, requirements for administration, and scoring. Further discussion and agreements will be needed.
- Is a cluster of measures being considered? Some discussed measures (such as observation checklists) would not be appropriate for all uses.
- How will items align with the new standards, including depth of knowledge? Selected standards only? At all possible DOK? Sets of items measuring multiple standards and/or multiple items measuring a standard?

**Whom to assess:** All schools, including the School for the Deaf and Blind. All students in the classroom, except those for whom the assessment would not be informative (e.g., students with severe cognitive disabilities).

*Visually impaired students:* need Braille and/or large print versions.

*ELL students:* In accordance with the consent decree (affecting only K-3, in this case), administration would be in English with some accommodations (e.g., question might be asked in Spanish, with the student responding in English). This means the test could be administered in one language and scored in another.

The consent decree requires translation into Spanish and Haitian Creole only for parent communications, not student assessments. However, we don't want language to be a barrier to assessing mathematics reasoning.

### **Type of Assessment:**

Need to clarify definitions and agree on the type of assessment desired:

Formative assessment - Generally defined as informal, administered frequently by the teacher in the classroom. Many instructional materials have embedded formative assessments with a range of quality.

Interim assessment - essentially "mini FCATs" - an intermittently administered summative assessment covering the full content domain over the course of a year. Mary Jane prefers not to define "when" the formative assessments should be given. If FDOE defines "when," it will be about getting kids to the same point, not about instruction. Teachers need some flexibility in when they teach various standards. If FLDOE does define "when" (e.g., every 6 weeks), it must be with the understanding that teachers will be teaching and assessing different content at any set point of time.

Progress monitoring: The purpose of progress monitoring is to identify underlying weaknesses, not determine whether students know the right or wrong answer. The true notion of progress monitoring may not fit mathematics, broadly defined, as it does reading. In reading, assessments measure growth or progress toward acquisition of a set of core skills (e.g., phonemic awareness, reading fluency). In mathematics, assessments measure progress in acquisition of generally discrete skills. Progress monitoring doesn't make sense for loosely related discrete skills.

### **Content of Assessment:**

State standards: The Next Generation standards in mathematics are more specific than the former standards and are specified for each grade level (not grade bands). They cover fewer concepts per grade level, with the expectation that concepts will be taught in greater depth.

Skills that predict future mathematics performance: Research on the predictive power of certain skills (e.g., phonemic awareness or fluency in reading) is lacking in mathematics. Ben Clarke has some curriculum-based measures, but most of his work relates to general math scores predicting later math scores.

Recommendation: Both of the above are important. The assessment will need to be standards and skills-based. Standards are important because the assessment will drive instruction. But the assessment should transcend curriculum. A curriculum-based assessment should not be the goal.

### Content framework:

- FDOE has an organizational scheme based on the content in the Next Generation K-3 standards and findings in the National Mathematics Advisory Panel report.
- The FCAT Specifications will be necessary to identify what students are expected to know and do at each grade level. FCR-STEM requested that FLDOE send the specs for grades 3-5 within the next month.

- Mary Jane and Todd drafted learning progressions across the grades. However, they're not researched-based. They were developed based on a "common sense" approach and comparisons of Florida's standards to those of Singapore and South Korea (looking at grades at which students are expected to learn and master different concepts and skills). These will be provided to FCR-STEM.

#### Possible item formats:

It is envisioned that administration will consist primarily of teacher and student interaction. Teachers would observe; students would build, draw, respond to questions. Manipulatives would be used as appropriate.

- Depth of understanding: Other countries build understanding of concepts well beyond procedural knowledge. They teach fewer concepts in greater depth (as is the intent of the Florida's new standards and benchmarks). How can in-depth understanding be assessed? One minute probes may be fine for reading but not what is needed in mathematics.
- Use of technology – Mary Jane would like to see students assessed using technology as soon as appropriate. Online administration expands possibilities for assessment. Val Shute (FSU) has done excellent work in this area. In general, research results for online assessment of young children (e.g., grade 3) are equivocal. Computer-based tests may or may not elicit student knowledge of content depending upon individual student experience with keyboard/mouse. DIBELS is not administered online. However, it's possible that technology could be used in some ways (e.g., point to the circle on a touch screen). Karen DeMeester mentioned a computer-based assessment tool (being used by Mike Spector at FSU) that captures student thinking. She will provide more information. Issue: Inadequate number and quality of computers in classrooms.
- Proof and defense of solutions will be expected across grades K-12. Younger children may be asked to draw a model, rather than write a response.
- Revealing student misconceptions (not just the right or wrong answer) is important.
- Reading should be incorporated.
- Example of item formats: The National Council on Teacher Quality's publication, *No Common Denominator: The Preparation of Elementary Teachers in Mathematics by America's Education Schools*, contains sample questions for teachers and students. Mary Jane suggested looking at sample items in the appendices in this publication (see [http://www.nctq.org/p/publications/docs/nctq\\_ttmath\\_fullreport\\_20080717081714.pdf](http://www.nctq.org/p/publications/docs/nctq_ttmath_fullreport_20080717081714.pdf).) She also noted that FLDOE is trying to change how teacher knowledge is assessed for certification. Laura recommended looking at the Learning Mathematics for Teaching (LMT) items at University of Michigan as well (see <http://sitemaker.umich.edu/lmt/home>).

#### Other comments:

- All of the benchmarks are rated on Depth of Knowledge (a modification of Norman Webb's system used in rating FCAT items), indicating the cognitive demand to be expected of students in instruction. With these new DOK ratings (1 through 3), the formative assessment will be measuring benchmarks at different levels.
- Conceptual understanding versus procedural knowledge should not be viewed as "either-or." The assessment will need to measure both. Let the data determine what items are most predictive of later performance. Item performance will also serve to validate learning progressions.
- Teachers often don't have the conceptual knowledge to understand students' answers or see where they are conceptually. Conceptual understanding should be taught before

procedural knowledge and algorithms. Formative assessments, if properly constructed, provide an opportunity to change teacher thinking about instruction.

**Mode of administration:**

It would be helpful to talk to FCRR about the DIBELS and new reading assessment roll-outs. Teachers assessing their own students on these reading measures at Reading First schools is new.

Preparation of teachers for test administration will depart from “business as usual.” They will need content and pedagogical content knowledge.

**Focus of literature review:**

- Core skills/knowledge in early grades mathematics – what we currently know
- Learning progressions and development of mathematical knowledge.
- Existing formative and summative measures in PK-3 mathematics – what we can learn or borrow
- Predictive validity studies – major findings and the analyses used
- Accountability burdens and requisite qualities of student assessments

UCF has extensive literature reviews on RTI with content foci on early literacy and mathematics. Mary and Shelby would be happy to share these and other resources.