Purpose and Item Selection of EMC-PK2

In 2015, COHERE PIs[1] outlined the need for a different kind of tool for observing Pre-K through 2nd grade math coherence.

> Existing measures of classroom quality have been used to assess structural elements of the classroom, classroom processes, or the quantity and quality of instruction in a particular subject matter area (e.g., literacy or math instruction). While each of these measures has its strengths, the field lacks a comprehensive, multidimensional assessment of structural elements, classroom processes, and the quality of instruction in different subject matter areas that are associated with child outcomes across the transition from preschool to early elementary grades. The Assessment Team will address this need, either by improving an existing tool or by developing a new one that is predictive of school achievement. The tool should have the following features:
>
> It should measure a range of structural elements and process features that support children's learning and achievement in preschool and early elementary school classrooms.
>
> It should address global and domains specific aspects of the classroom. For example, global features include overall quality of the classroom environment, including structural elements and process features; domain-specific features include the quality of interactions or practices in a specific areas, such as math, social emotional development, or instruction. (IES proposal, 2015)

The COHERE research group further defined dimensions of what was called *Coherence-3*:

• Subject-matter (or domain) coherence refers to the degree to which presentations of subject matter content accurately embody the discipline; for example, concepts, facts, relationships, and processes are in line with those of domain experts, clearly represented and/or explicated, and interconnected.

• Psychological coherence refers to the degree of consistency of the classroom environment, curriculum, and teaching with patterns of student thinking and learning in ways that support students' meaningful engagement in the subject matter and continued development in understandings and skill.

• Instructional coherence refers to the degree to which teaching activities and strategies are consistent with research on effective instruction and tasks; for example, the extent to which teachers use instructional tasks and strategies that help students connect and relate different experiences, concepts, and representations of concepts.

[1] Doug Clement, Cynthia Coburn, Dale Farran, Megan Franke, Deborah Stipek (later added Kelley Durkin)

How the items were selected

In June, the Vanderbilt University team[2] began a review of existing classroom observation measures. The search resulted in 55 different measures. Of those measures, the team obtained 38 complete measures. The complete measures that were unavailable during this pass required either purchasing from the developer, or required completing a complete training course, or were not published, or were simply not readily available. However, some of their elements were described or referenced in the literature. An analysis of the observation measures coded existing measures along the following dimensions:

	Intended for PK-2 (inclusive)	Included math-specific items
N Systems that Met Criteria (Total N=55)	8/55	31/55

Length of observation	Less than 1 hour/ lesson	1 hour/lesson	Partial school day	Full school day	Multiple lessons or days	Not specified or unknown
N Systems that Met Criteria (Total N=55)	7	21	9	2	3	13

Ratings-based or Narrative/Notes-based	Ratings-based	Narrative/ Notes-based	Both	Other	Unknown
N Systems that Met Criteria (Total N=55)	33	3	7	9	3

Environment/Practices/ Interactions Based	Environ-ment	Practices	Inter-actions	All	Combin-ations of 2 cate-gories	Unknown
N Systems that Met Criteria (Total N=55)	1	14	3	12	17	8

Teacher or Student Focused	Teacher-focused	Student-focused	Both	Unknown
N Systems that Met Criteria (Total N=55)	15	3	31	6

[2] Dale Farran, Sascha Mowrey, Luke Rainey

While many existing measures were math-focused (at least in part), very few were designed for use in both Pre-K and early elementary grades. A plurality of measures was meant for the length of a lesson – usually a partial school day in Pre-K and 1 hour in elementary and above. Some measures were designed to work only with recorded video from lessons. Most measures involved some type of ratings, though some combined ratings and notes. Most measures focused on the quality or frequency of specific classroom practices, though many looked at combinations of different domains like the classroom environment and teacher-student interactions. Most measures were both teacher and student-focused, but many others were only focused solely on teacher practices or behavior.

The observation team determined that both the COEMET (Sarama & Clements, 2007) and the Advanced Narrative (Farran et al., 2015) were the two instruments most relevant to the goals of this project. Each was designed for live observations, was designed with Pre-K in mind, included both math and global components, focused on environment, practices, and interactions, and focused on both teacher and student behaviors. The COEMET structure allows for a closer look at the activity level but the Advanced Narrative provides a better overall perspective. Together, the measures offered valuable templates for collecting both qualitative and quantitative data across a math lesson. The VU team decided to adapt conceptual and structural components from both measures into a new observation measure.

In determining which components to use from these tools, the observation team cross-walked them with the elements of *Coherence-3*. This offered clarity into where new items might be needed that neither tool featured. Through weekly discussion group meetings and in-classroom pilot observations, the team developed both structure and content that met the conceptual and practical needs. To round out the environmental rating items, additional ratings were incorporated from the Post Observation Rating Scale (Farran et al, 2014).

The first version of the new tool, called the Early Math Coherence Pre-K through 2nd Grade (EMC-PK2) measure, was used to collect data during the first cross-sectional year of the COHERE project. Observing teachers at each grade, Pre-K through 2nd grade, across six schools in two California districts, the team gathered valuable information about the validity of the tool. After the cross-sectional study, the observation team revised and made improvements to the tool for the upcoming longitudinal study. This final version was used all four years of the study without major revisions.

The components of the tool are listed below, along with their origins.

Cover Page

Number of Students Present

This was adapted from the cover page of the Advanced Narrative and the COEMET.

Number of Parents/Volunteers

This was adapted from the cover page of the Advanced Narrative and the COEMET.

Number of Other Staff

This was adapted from the cover page of the Advanced Narrative and the COEMET.

School Name, Lead Teacher ID, Lead Teacher Name and **Grade Level**

This was adapted from the cover page of the Advanced Narrative and the COEMET.

Teaching Assistant Present:

This was adapted from the cover page of the Advanced Narrative and the COEMET

Observation Start Time and **Observation End Time**

This was adapted from the cover page of the Advanced Narrative and the COEMET. The Advanced Narrative recorded the School start and end time on its cover page, but included timestamp fields elsewhere in the measure.

Total Full and Mini IMAs

IMAs were adapted from the COEMET's SMAs (Specific Math Activities). Like IMAs, SMAs were records detailing each math activity during a lesson. A single observation would likely have multiple records completed. Mini and Full IMAs were also categories of SMAs in the COEMET, where Fulls involved significant teacher interaction and Minis did not. For more information about these distinctions, and how IMAs differed from SMAs, see p. 6. This field in the EMC-PK2 automatically tallied the total number of each type of IMA on the Cover page of the measure.

Cover Notes:

Adapted from the Advanced Narrative, this notes field on the EMC-PK2 cover was changed from being a place to record significant events/behavior issues to being a place to record general notes about the observation. Observers were instructed to explain their reasoning here for their ratings in the Post-Observation section.

Intentional Math Activities (IMAs)

IMA Start and End Time

These were originally tied to an Advanced Narrative structure in the cross-sectional version of the tool, which recorded episodes of time across the observation, and extending to non-math subjects. For the longitudinal version of the tool, the episode structure was removed and observers only focused on math episodes during IMAs. To understand how much transition and non-math time occurred during an observation, analysts could take the total time in IMAs and subtract it from the overall observation time, recorded on the Cover page.

Repeat of IMA #

This item was added for the longitudinal iteration. The concept was adapted from the COEMET, which provided fields to record different rotations of an IMA, and how many students participated each time. After pilot testing, the observation team determined that this system made it difficult to compare how teachers might adapt their interactions for different groups of students. The new longitudinal system involves a decision rule, which requires observers to create new IMAs for rotations of new students. They use this field to identify the original IMA number for which this is a rotation.

IMA led by

This was an item in the COEMET, removed during the cross-sectional year but added back in for the longitudinal version. During the pilot observations, there was some interest in comparing the effectiveness of IMAs led by the lead teacher compared with other classroom staff.

Launch

This was an item added after the cross-sectional year. In one of the early drafts of the measure, there was a binary item for determining whether an IMA objective was clear or unclear. However, it was left out of the cross-sectional version. While piloting, a member of the research team noted that in several cross-sectional classrooms, their observations were quite difficult to interpret because the lesson objectives were never made explicit. For example, students would come into the classroom and begin working on a problem on the board, then start on a workbook page with no teacher direction. In other classrooms, the lesson objective was written on the board, and the teacher regularly reminded student how their tasks related back to the objective. During the cross-sectional year, it was commonplace in the earlier grades for students to select their own activities in centers, and the math objective could only be implied by observing the task. This item was created to highlight potential classroom and grade similarities and differences in how tasks were launched.

Language(s)

Several classrooms were designated Spanish-language immersion programs across all subjects. However, observers discovered some other classrooms not designated immersion programs

included a good deal of bilingual Spanish and English instruction. We also discovered some classrooms with non-Spanish immersion (such as a Mandarin immersion Pre-K) and bilingual students speaking other languages than Spanish. For the longitudinal version of the measure, the team created a checklist on the IMA Cover page to track the language use at the IMA level. In some cases, the observer was not fluent in the language and may have not understood some interactions, so the team wanted to flag those activities. It also provided a record of differentiation and accommodation practices in certain classrooms.

Activity Type

Both the COEMET and Advanced Narrative had space to record the activity type (e.g. Whole Group, Small Group, etc.). In the Advanced Narrative, each episode with at least 75% of class participation was coded for Activity type, but the COEMET only recorded it for each math activity. During the cross-sectional study year, the EMC-PK2 included an Advanced Narrative section in which each activity, whether math or non-math, received an activity type code. This was dropped in the longitudinal version of the EMC-PK2, which only coded math activities. Many of the codes are the same as the Advanced Narrative, but some were consolidated (for instance, Small Group Teacher Center and Small Group Teacher).

Auxiliary Activity Type

After the cross-sectional year, the observation team consolidated Activity codes to focus only on ones that would occur during math activities. However, there were some instances during the cross-sectional study, in which we observed math embedded in other activities. We created the Auxiliary type codes to record these special cases more systematically. We adapted the Advanced Narrative's Meal Time (Meal) and Playground, Special (Outside the Classroom) codes, and added Choice, to differentiate chosen Centers from Teacher-directed small groups.

Content Connections

This section was originally adapted from Full SMA rating scales in the COEMET and in the Advanced Narrative. They were first broken out as a checklist during the cross-sectional observation year.

Student Practices

The Common Core lists eight Standards for Mathematical Practice derived from the NCTM and the National Research Council. These are described as important ways students should engage in math, to accompany the standards describing what kind of math content they should be engaged in. During the cross-sectional year, the observation team used this list to create a checklist of 14 student math practices observers might see across the entire lesson. Observers were to mark whether each practice happened, and how often. However, this system proved too difficult to

reliably observe, while simultaneously observing other global and teacher-based practices. Additionally, the cross-sectional year data showed that observers rarely encountered most practices in the list once, let alone multiple times. For the longitudinal study, the team revised the tool to focus on five consolidated practice categories. We also moved this checklist to the IMA level and removed the frequency rating. Observers could select whether the practice occurred at least once during each IMA or not.

Worksheets

This item was added to the original cross-sectional year version. It was adapted from a code for worksheets in the Advanced Narrative. There was also a rating item in the Post observation rating scale which was adapted.

Full/Mini IMA

Adapted from the COEMET's Specific Math Activity (SMA), the Intentional Math Activity (IMA) was also a record of a math activity during a lesson. It retained the designation Full and Mini from the COEMET, with Fulls generally describing activities with significant teacher involvement, and Minis involving very little teacher involvement. The decision rule for determining Full/Mini in the EMC-PK2 is specifically defined as having greater than or less than a minute of consecutive teacher involvement in an activity, beyond giving behavioral instructions.

IMA Notes

The notes component of the tool derives mainly from the Advanced Narrative tool. In the cross-sectional study year, observers recorded notes during each episode of the lesson, whether math or another subject. For the longitudinal tool, the main note taking was moved to the first page of each IMA record. Observers were asked to spend most of the observation recording detailed notes, and then use those notes to complete their coding. This process allowed for a rigorous data checking process to take place after each observation, where another analyst could match the codes with the notetaking to confirm the validity of the coding. The notes also allowed analysts to code each IMA's focal math content and provided important context for later analysis of the data.

Math Content

Both the COEMET and the Advanced Narrative included items recording the focal math content of the observation. The EMC-PK2 includes a broad selection of potential math content, divided into Domains (e.g., Counting and Cardinality) and Subdomains (Counting 1-1 within 5). It includes items corresponding with standards from Pre-K through 3rd grade. Members of the

COHERE team recommended including 3rd grade content in case some 2nd grade classrooms were observed teaching advanced content. The items are categorized by Domain in the tool, but the observer does not know which item corresponds with which grade level.

The original intent was for observers to become reliable coders of math content. However, it became clear during the cross-sectional year that the 90+ content codes possibilities would be very difficult for observers to become reliable in a reasonable timeframe. Because of how the standards are written, there are many content codes that could apply to different practices and require subtle distinctions in the coding scheme. In addition, it is sometimes difficult to discern whether a particular math content is "focal" to the lesson objective, or whether it is simply a building block of the lesson (e.g., counting while solving an operations problem). As a solution, we trained research staff to double-code each IMA using the observer notes, according to a detailed coding scheme. Coding discrepancies were resolved through conversation between the coders, and decision rules were saved in a shared crosswalk document.

Originally, the codes were taken from the California Pre-K Foundations and the Common Core Math Standards. They were adapted into observable descriptive items, since many standards were written in general terms, or overlapped across multiple grade levels. After discussions with the research group, we cross-walked and coded each subcontent item to its corresponding grade level. Each IMA could receive an average math content score, to determine whether it was focused on, above, or below grade level. Several items were not averaged in since they were written across multiple grades in the standards (e.g., using word problems). Some additional items were added that do not correspond with standards, but often appear during designated math periods (e.g., naming days or months of the year).

IMA Ratings

This section was adapted from the COEMET's SMA Ratings section. However, individual items derived from both Advanced Narrative and COEMET. Many items from the COEMET were rated on a Likert scale or were yes/no questions, so the research team converted them into ratings scales with behavioral anchors. All of the ratings were iterated through pilot testing. Several were added and dropped after the cross-sectional year due to lack of clarity or redundancy.

IMA Rating 1 –

Both Advanced Narrative and COEMET included items about responsiveness to student contributions.

IMA Rating 2-

Some language about scaffolding using incorrect responses derives from the Advanced Narrative. The behavioral anchors also relate to a scaffolding paper by Anghileri (2006).

IMA Rating 3 –

Both Advanced Narrative and COEMET included items about question asking. After the cross-sectional year, edits were made to clarify the anchors to this item. The revision shifted the focus of the question from the frequency of questioning to the quality of questioning.

IMA Rating 4 –

This item was added for the longitudinal version of the tool from the work of Smith & Stein (1998) on cognitive demand. The cross-sectional version of the tool has ratings focused on the "appropriateness" of the task, but that proved to be at time a nebulous construct for observers.

IMA Rating 5 –

This item derives from the Advanced Narrative and Post observation rating scale. After the cross-sectional year, the intent of this item was clarified to focus in on student participation.

IMA Rating 6-

Some scaffolding language at the activity level derives from the COEMET. However, this item was adapted by the research group to focus on individualized adaptations as a particular type of scaffold.

IMA Rating 7 –

The language for student engagement behavioral anchors comes from the Advanced Narrative.

Post-Observation Ratings

Most of the items were adapted from the Advanced Narrative or The Post Observation Rating Scale (PRS).

Post 1-

This item derived from COHERE research group recommendation, although some elements relate to a rating in the Advanced Narrative.

Post 2 –

This item derived from COHERE research group recommendation, although some elements relate to a rating in the Advanced Narrative.

Post 3 –

This item was adapted from the Advanced Narrative.

Post 4-

This expanded on a rating from the Advanced Narrative focusing on social supports in the classroom environment.

Post 5 –

This item was adapted from a COEMET item about connections.

Post 6-

Elements in this item originated from the Advanced Narrative and Post observation rating scale

Post 7 –

This item derives from the Post observation rating scale

Post 8-

This item derives from the Post observation rating scale

Math Environment Materials

These were adapted from a checklist in the COEMET.

Red Flags

These were adapted from items in the Post observation rating scale.

The following is a document created by the development team at VU in 2015, describing the early version of the observation tool for the larger COHERE team.

COHERE Classroom Observation System

Changes Effective November 19, 2015

OVERVIEW

Contributors include Dale Farran, Sascha Mowrey, Luke Rainey, Abigail Do, Laura Piestrzynski, Nicholas Kochmanski

1. Definition of SMA

The operational definition of a Specific Math Activity will be an activity with math as its focus that for the duration of the activity does not change in either its content focus or its pedagogy. If either the math content being taught or the process by which it is taught changes, we will begin a new SMA.

In the K-2 classrooms we observed, especially 1st and 2nd grades, the math lesson of the day tended to be longer than we see in pre-k often for the purpose of providing different activities to give children a chance to practice or to experience the same math content in a different way. Rather than calling that a single, long SMA, we will start a new SMA when the way the content is taught changes (e.g., from a whole group lesson to seat work). Our rationale is that we saw very different teacher behaviors when the context changed. Rather than average our ratings across contexts, we will provide the SMA ratings for the math content taught in each specific context, in other words as a new SMA.

2. Math Content – Crosswalk between Common Core, SFUSD and California.

We have done the cross walk to examine the connections between the "Content Emphases" in SFUSD and CA and the Common Core Standards. We were relieved to see how related they are (attached). The decision we are recommending is to go with the Common Core Standards for our content codes in the observation system, the way we have it now. We can, however, in the analyses report back to SFUSD with its own terminology showing the connections. As we work with other systems in other states, we should conduct this cross walk again. If we encounter a system whose content emphases have discrepancies from the Common Core, we can always add new areas to the observation system. By making the Common Core standard our anchor, we will have something consistent across school districts in different states.

3. **Coherence-3 Dimensions**

We are very grateful to Doug for steering us in the direction of the Coherence-3 dimensions described in the proposal to IES. They have served to reorganize our thinking in major ways. We have gone through the observational items – including those for the SMAs and those done at the end on the POST – and grouped them into the appropriate domain. We view the domains in short hand and roughly to be about the "content" (Subject-Matter coherence), "student learning – or child level" (Psychological coherence) and "instruction—or teacher level" (Instructional coherence). Doing so revealed several things. First, the Psychological area was weak. We have addressed that by adding a section to the Post on Practices that I will describe in another section. Our survey also showed items that needed to be clarified in order to be clear to which area they belonged. You will see that we added a fourth dimension that we labeled Moderators, and that will be described below.

Moreover, we have reorganized the SMA items and the POST so that all items related to one dimension are grouped together. I am attaching the new version with the changed items all rearranged into the Coherence-3 dimensions and highlighted by dimension. We believe that this grouping will make it much easier for people to learn and to carry out the system, so important for training as well as implementation.

4. Mathematical Practices Observed Among Students Across the Day

We wanted to follow the COHERE faculty's suggestions that we try to include strategy use by students. These had to be observable behaviors. Sascha and Luke went back through the Common Core Standards for Mathematical Practice. They isolated key details and figured out observable student behaviors that reflect the practices. The entire Research Group reviewed the strategies to determine any ones missing and whether these could indeed have been observed when they were in classrooms.

The result is a **new** section of the POST. It comes at the end and consists of a list of 14 student behaviors. Observers will mark whether these behaviors were observed frequently, a few times, or never. In order to do this, observers will be trained to make notes in the Narrative Record when they see any of these behaviors (easier to do than you might at first think). They can review their written notes before marking this section. We debated making this section simply a Yes-No (saw it or didn't) but decided that observers could likely tell if they saw it a lot, a little or not at all without great difficulty.

5. **Moderators**

You will see that we have defined these in the Coherence-3 document as aspects of the classroom that could prevent or facilitate all of the three coherence domains. They relate to things like how smoothly the classroom ran, what the teacher's tone was like, etc. We have marked all the items related to this dimension in the text you are getting.

6. **Red Flags**

At the very end of the POST is a list of 6 things that if an observer sees it even once, he or she notes.

7. **New Items and old COEMET Items**

We very carefully reviewed (several times) the items Doug sent from the old COEMET and that Deborah very helpfully provided some statistical information about. All except one seemed to be covered by items we already had. Our items might look a bit different because we used anchors instead of a Likert scale, but we believe that the information in the old items is included in the anchors of the new ones. POST item #4 is new. We are attempting to address the differentiation among students that Deborah suggested. We have another item that relates to this same area (#5). We see the distinction being that #5 is about adapting the math **content** to student's ability to understand and work with it whereas #4 is broader – do teachers recognize and adapt to different needs of their students (e.g., one student might need a quiet place to work, one might need manipulatives, another might need help getting an activity going – the math content might stay the same but the task is altered to fit the needs of the students). We think observers can see both of these things and distinguish them.

References

Anghileri, J. (2006). Scaffolding practices that enhance mathematics learning. J Math Teacher Educ 9, 33–52. https://doi.org/10.1007/s10857-006-9005-9

Child Development Division, California Department of Education. (2008b). California Preschool Learning Foundations (Volume 1). California Dept of Education. https://www.cde.ca.gov/sp/cd/re/psfoundations.asp#psfoundvol1

Farran, D.C., Meador, D. M., Keene, A., Bilbrey, C., & Vorhaus, E. (2015). Advanced narrative record manual, 2015 Edition. Nashville, TN: Vanderbilt University, Peabody Research Institute.

Farran, D.C., Meador, D. M.,& Yun, C. (2014). Post observation rating scale manual, 2015 Edition, adapted for Preschool Expansion Classrooms. Nashville, TN: Vanderbilt University, Peabody Research Institute.

Institute for Education Sciences (2015). *Supporting Early Learning from Preschool Through Early Elementary School Grades: Proposal for an Early Learning Assessment Team.* Retrieved from https://ies.ed.gov/funding/doc/2016_84305N.docx

National Governors Association Center for Best Practices, Council of Chief State School Officers. (2010). Common core state standards: mathematics. Washington D.C. http://corestandards.org/

Sarama, J., & Clements, D. H. (2007). Manual for Classroom Observation (COEMET) — Version 3. Unpublished Version

Smith, M., & Stein, M. (1998). Reflections on practice: Selecting and creating mathematical tasks: From Research to Practice. Mathematics Teaching in the Middle School, 3(5), 344-350.

Yun, C., Farran, D., Lipsey, M., Vorhaus, E., Meador, D. (2010). Observer Manual for the Post Observation Rating Scale for the Tools of the Mind Curriculum (PRS – ToM). Retrieved from https://cdn.vanderbilt.edu/vu-my/wp-content/uploads/sites/412/2012/01/14084339/Post-Observation-Rating-Scale_Manual_061610_web1.pdf