**COHERE Observation Subscale Development**

**Observation Overview**

The Early Math Coherence PK to Grade 2 (EMC-PK2) observation system was developed over a 19-month period to capture the coherence and alignment of mathematics instruction from PK through 2nd grade. The observation system results in data on many variables, and this document describes a set of subscales that can be used for data analysis. It includes details on why these subscales were chosen and how they were created.

The data that inform the analyses below come from observations that occurred during the DREME COHERE longitudinal study. Classrooms were observed once in the fall, once in the winter, and once in the spring. However, due to the pandemic, there were no spring observations in 2020, when most students were in 2nd grade. These observations occurred during the designated math instructional period, except for PK where the observations occurred over a 3-hour period. The focus of the observations was on the content of the math being taught, the promotion by the teacher of reasoning and problem solving, pedagogical strategies used by the teacher, involvement of the children in deeper thinking about math, and the structural coherence of the math lessons.

**IMA Ratings**

Intentional Math Activities (IMA) represent discrete episodes of math instruction. Full IMAs feature direct math instruction and/or interaction between teachers and students, and Full IMAs are rated on seven items with behaviorally anchored 1-5 rating scales. (Mini IMAs on the other hand have students working on math activities without direct teacher interaction and are not rated on the seven items.) These seven IMA ratings have been used to develop two subscales: **teacher responsiveness** and **student engagement**. Additionally, IMA rating 6 has been recoded into a binary variable due to its lower frequency.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Descriptive Statistics** |  | |  |  | |
|  | N | Mean | | | SD |
| 1. Teacher listens and responds | 960 | 2.33 | | | 1.08 |
| 2. Teacher utilizes incorrect responses (only rated when incorrect responses given) | 543 | 2.74 | | | 0.84 |
| 3. Teacher asks questions | 960 | 2.27 | | | 1.04 |
| 4. Teacher maintains cognitive demand | 960 | 2.35 | | | 0.90 |
| 5. Behavioral and participation instructions are clear and sufficient for student success | 960 | 4.39 | | | 0.82 |
| 6. Teacher adapts the task | 960 | 1.42 | | | 0.92 |
| 7. The engagement level of the groups of students involved in this IMA | 960 | 3.65 | | | 0.94 |

|  |  |  |
| --- | --- | --- |
| **Task Adaptation - Binary** |  |  |
|  | Count IMAs w/ Adaptation (N=960) | Percent IMAs w/Adaptation (N=960) |
| Task Adaptation | 212 | 22.1 |

Correlations between items were calculated to determine which groupings of items might be candidates for subscales. The table below shows the strongest correlations between Items 1, 2, 3, and 4 and items 5 and 7. Item 6 did not correlate strongly with any of the other items.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Correlations** | |  | |  | |  |  | |  | |
|  |  | | 2. Incorrect responses | | 3. Asks questions | | | 4. Cog demand | | 5. Participation | | 6. Adapts the task | 7. Engagement |
| 1. Listens & responds | *r* | | .500\*\* | | .670\*\* | | | .554\*\* | | .088\*\* | | .213\*\* | .124\*\* |
| *p* | | <.0001 | | <.0001 | | | <.0001 | | .006 | | <.0001 | <.0001 |
| 2. Incorrect responses | *r* | |  | | .426\*\* | | | .450\*\* | | .177\*\* | | .189\*\* | .254\*\* |
| *p* | |  | | <.0001 | | | <.0001 | | <.0001 | | <.0001 | <.0001 |
| 3. Asks  questions | *r* | |  | |  | | | .628\*\* | | .093\*\* | | .115\*\* | .102\*\* |
| *p* | |  | |  | | | <.0001 | | .004 | | <.0001 | .002 |
| 4. Cognitive demand | *r* | |  | |  | | |  | | .052 | | .133\*\* | .133\*\* |
| *p* | |  | |  | | |  | | .111 | | <.0001 | <.0001 |
| 5. Participation | *r* | |  | |  | | |  | |  | | .059 | .596\*\* |
| *p* | |  | |  | | |  | |  | | .068 | <.0001 |
| 6. Adapts  the task | *r* | |  | |  | | |  | |  | |  | .080\* |
| *p* | |  | |  | | |  | |  | |  | .013 |

\*\* *p* < .01, \* *p* < .05

**Factor Analysis**

Factor analysis was used to further inform the composition of IMA subscales. Principal Axis Factoring with Varimax rotation was used with pairwise deletion of missing cases due to the nature of IMA rating 2. A sensitivity check using listwise deletion for missing data was also performed with very similar results. The Rotated Factor Matrix for the pairwise deletion analysis is included here, and full results from both analyses are available upon request. The same pattern that emerged in the correlations was found in the factor analysis results.

|  |  |  |
| --- | --- | --- |
| **Rotated Factor Matrixa** | | |
|  | Teacher Responsiveness | Student Engagement |
| 1. Listens & responds | .811 | .041 |
| 2. Incorrect responses | .582 | .212 |
| 3. Asks questions | .807 | .013 |
| 4. Cognitive demand | .733 | .031 |
| 5. Participation | .069 | .704 |
| 6. Adapts the task | .209 | .074 |
| 7. Engagement | .122 | .832 |
| Extraction Method: Principal Axis Factoring. | | |
| Rotation Method: Varimax with Kaiser Normalization. | | |
| a Rotation converged in 3 iterations. | | |

After confirming that the candidate items were conceptually related, Cronbach’s alpha was calculated for each grouping of ratings for the proposed subscales:

IMA Subscale 1: Teacher Responsiveness

Cronbach’s α: .823

IMA Subscale 2: Student Engagement

Cronbach’s α: .743

**POST Ratings**

After completing an observation, observers rate their observations on eight items with a 1-5 scale using behavioral anchors. These items were used to develop two subscales: **differentiation** and **classroom atmosphere**.

|  |  |  |  |
| --- | --- | --- | --- |
| **Descriptive Statistics** | | |  |
|  | N | Mean | SD |
| 1. Activities were linked by a non-math theme or “storyline” | 279 | 1.27 | 0.80 |
| 2. Activities were linked by a “big math idea” or math concept | 279 | 3.44 | 1.32 |
| 3. How the teacher reinforced math learning | 279 | 2.35 | 1.04 |
| 4. Across the observation the teacher accommodated the range of student abilities and development to promote individual students’ thinking and learning | 279 | 2.03 | 1.03 |
| 5. The teacher communicated math concepts in multiple ways | 279 | 2.07 | 1.10 |
| 6. The atmosphere of the classroom environment is positive | 279 | 3.88 | 1.08 |
| 7. Behavior management during lessons/activities is minimal | 279 | 3.65 | 1.20 |
| 8. The tone of the teacher’s interactions with students is positive | 279 | 3.78 | 0.90 |

Correlations between items were calculated to determine which groupings of items may be candidates for subscales. The table below shows the strongest correlations between Items 3, 4, and 5 and items 6, 7, and 8. Items 1 and 2 did not correlate strongly with any of the other items.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Correlations** | | | | | | | | | |
|  |  | 2. Big math idea | 3. Reinforce learning | | 4. Accommodates | 5. Multiple ways | 6. Class environ | 7. Behav manage | 8. Teacher tone |
| 1. Non-math theme | *r* | .086 | .354\*\* | | .151\* | .261\*\* | .132\* | .081 | .130\* |
| *p* | .151 | 0 | | .012 | 0 | .028 | .175 | .03 |
| 2. Big math idea | *r* |  | .248\*\* | | .106 | .243\*\* | .089 | .065 | .135\* |
| *p* |  | 0 | | .078 | 0 | .137 | .277 | .024 |
| 3. Reinforced learning | *r* |  |  | | .354\*\* | .534\*\* | .365\*\* | .252\*\* | .369\*\* |
| *p* |  |  | | 0 | 0 | 0 | 0 | 0 |
| 4. Accommodates | *r* |  |  | |  | .466\*\* | .243\*\* | .091 | .222\*\* |
| *p* |  |  | |  | 0 | 0 | .131 | 0 |
| 5. Multiple ways | *r* |  |  | |  |  | .283\*\* | .123\* | .290\*\* |
| *p* |  |  | |  |  | 0 | .04 | 0 |
| 6. Class environment | *r* |  |  | |  |  |  | .657\*\* | .675\*\* |
| *p* |  |  | |  |  |  | 0 | 0 |
| 7. Behavior management | *r* |  |  | |  |  |  |  | .409\*\* |
| *p* |  |  | |  |  |  |  | 0 |
| \*\* *p* < .01, \* *p* < .05 | | | |

**Factor Analysis**

Factor analysis was further used to inform the composition of POST subscales. The Principal Axis Factoring extraction method was initially attempted; however, extraction was unsuccessful, seemingly due to multicollinearity issues between Items 6 and 8. As a result, the Generalized Least Squares method was used instead. Varimax rotation was used to create the rotated factor matrix, which is included here. Full results are available upon request. The same pattern that emerged in the correlations was found in the factor analysis results.

|  |  |  |
| --- | --- | --- |
| **Rotated Factor Matrixa** | | |
|  | Classroom Atmosphere | Differentiation |
| 1. Non-math theme | .067 | .380 |
| 2. Big math idea | .036 | .308 |
| 3. Reinforced learning | .248 | .693 |
| 4. Accommodates | .154 | .521 |
| 5. Multiple ways | .154 | .755 |
| 6. Class environment | .984 | .174 |
| 7. Behavior management | .656 | .061 |
| 8. Teacher tone | .638 | .273 |
| Extraction Method: Generalized Least Squares. | | |
| Rotation Method: Varimax with Kaiser Normalization. | | |
| a Rotation converged in 3 iterations. | | |

After confirming that the candidate items were conceptually related, Cronbach’s alpha was calculated for each grouping of ratings for the proposed subscales:

POST Subscale 1: Differentiation

Cronbach’s α: .713

POST Subscale 2: Classroom Atmosphere

Cronbach’s α: .799

**Note**: POST Items 1 and 2 are being used as separate items due to their low correlations and poor factor loadings with the factors listed above.

**Student Practices**

Observers track whether students engage in certain practices related to math instruction during each IMA. These practices are not observed on a rating scale but are instead simply coded Yes/No depending on whether the observer witnessed the practice in each IMA. There are five practices in total, one of which is kept separate and four of which are combined into what are known as **Student Discussion Practices**.

|  |  |  |
| --- | --- | --- |
| **Frequencies of Student Practices – IMA level (N=1754)** |  |  |
| Practice | Count IMAs w/ Practice | Percent IMAs w/ Practice |
| Students used tools/visuals other than paper/pencil | 995 | 56.7% |
| Student Discussion Practices (talk, reflect, explain, steps) | 508 | 29.0% |
| Students talked with each other about math | 272 | 15.5% |
| Students reflected on math ideas | 40 | 2.3% |
| Students explained their reasoning | 172 | 9.8% |
| Students described steps they used to solve a problem | 185 | 10.5% |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Proportions of IMAS with Student Practices – Observation level (N=279)** | | | |  | |  |
| Practice | Min | Max | Mean | | SD | |
| Students used tools/visuals other than paper/pencil | 0 | 1.00 | .54 | | .33 | |
| Student Discussion Practices (talk, reflect, explain, steps) | 0 | .67 | .12 | | .12 | |
| Students talked with each other about math | 0 | .88 | .17 | | .22 | |
| Students reflected on math ideas | 0 | .50 | .03 | | .10 | |
| Students explained their reasoning | 0 | 1.00 | .13 | | .22 | |
| Students described steps they used to solve a problem | 0 | 1.00 | .13 | | .21 | |

**Investigation of Potential Rater Biases**

To investigate whether any of the observation items that use a rating scale were affected by rater bias, interclass correlations were calculated for IMA and POST ratings using covariance estimates from HLM models with Observer as a random effect. Low ICCs indicate a low likelihood of rater bias.

|  |  |
| --- | --- |
| **Measure** | **ICC** |
| IMA Rating |  |
| 1. Teacher listens and responds | .066 |
| 2. Teacher utilizes incorrect responses | .043 |
| 3. Teacher asks questions | .034 |
| 4. Teacher maintains cognitive demand | .030 |
| 5. Instructions are clear and sufficient | .068 |
| 6. Teacher adapts the task | .058 |
| 7. Student engagement | .234 |
|  |  |
| POST Item |  |
| 1. Non-math theme | .000 |
| 2. Big math idea | .195 |
| 3. Teacher reinforced math learning | .019 |
| 4. Teacher accommodated range of abilities | .137 |
| 5. Teacher communicated in multiple ways | .148 |
| 6. Classroom environment | .186 |
| 7. Behavior management | .181 |
| 8. Teacher tone | .087 |

Overall, the ICCs were very low, indicating a low likelihood of rater bias. The item with the highest ICC was the IMA rating of student engagement, but even this ICC was relatively low.