EXAMINING ELEMENTARY AND MIDDLE SCHOOL MATHEMATICS INSTRUCTION: ARE WE PROMOTING EQUITY AND ACCESS?

Anne Garrison Wilhelm
Southern Methodist University
awilhelm@smu.edu

Elizabeth Adams
Southern Methodist University
eladams@smu.edu

Jonee Wilson
North Carolina State University
Jwilson9@ncsu.edu

Temple Walkowiak
North Carolina State University
tawalkow@ncsu.edu

This descriptive study attended to the extent to which we see evidence of the presence of four practices that promote equity and access in 141 grades 3-8 mathematics lessons in the United States. We found that lessons generally showed evidence of some incorporation of the practices but often not at the highest level. Teachers in this sample engaged in social coaching at a relatively high level, across elementary and middle school classrooms. Teachers tended to do less with respect to supporting connection and engagement between student context and the math learning environment. We also found statistically significant differences between elementary and middle school lessons in positioning students as competent and supporting a nurturing environment by proactively building relationships and productive classroom culture. We offer possible interpretations and a few brief implications of these findings.

Keywords: Equity, Inclusion, and Diversity; Elementary School Education; Middle School Education; Instructional Activities and Practices

Instruction that meets national standards for student learning (National Council of Teachers of Mathematics, 2000; National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010) is considered "ambitious" because it is more demanding and requires more of teachers than the prior focus on procedural competence (Lampert et al., 2010; Lampert & Graziani, 2009). However, ambitious mathematics instruction remains uncommon and opportunities for students to develop the understandings outlined in the *Standards* are not distributed equally (Banilower et al., 2018; Boston & Wilhelm, 2017). This is particularly true for students of color and students for whom English is not their first language (e.g., Nasir & Cobb, 2002; Gutstein & Peterson, 2005). Recently, there has been progress towards identifying specific instructional practices that support historically marginalized groups of students, particularly as they participate in more rigorous mathematics. In order for teachers to develop and implement the identified practices, they need support in understanding distinctions between strong and weak examples of the practices (Goodwin, 1994; Grossman et al., 2013; Little 2003).

The Equity and Access Rubrics for Mathematics Instruction

Gutiérrez (2012) specified four dimensions to attend to the extent to which learning environments support historically marginalized groups of students and might be characterized as aiming for equity: access, achievement, identity, and power. Access and achievement comprise the "*dominant axis*" of equity, while identity and power make up what Gutiérrez called the "*critical axis*" of equity. The EAR-MI (Equity and Access Rubrics for Mathematics Instruction) is a set of classroom observation rubrics developed to capture seven practices that support marginalized students particularly along the dominant axis in gaining access to and more equitably participating in rigorous mathematics activity (Wilson et al., 2019). The EAR-MI was

designed by carefully observing the practices of middle school teachers in classrooms characterized by "*Standards*-based" instruction where historically marginalized students have been successful (Wilson, accepted). In this paper, we examine the classroom instruction of elementary and middle grades teachers to investigate the extent to which we see evidence of the presence of practices from four of the rubrics that produced reliable results and remained unchanged across two generalizability studies of the EAR-MI. Specifically, we ask the following question: To what extent are grades 3-8 teachers using some of the practices that support equity and access outlined in the EAR-MI?

Building an Argument for Validity

One intended use of the EAR-MI is to support researchers to assess teachers' progress in enacting mathematics teaching practices that support equity and access. Since the initial empirical study that described the identified classroom practices (Wilson et al., 2019), a set of rubrics were developed and have undergone numerous revisions. The rubrics were revised based on feedback from experts in ambitious and equitable mathematics teaching as well as experts in rubric development. Additional revisions were made based on multiple rounds of pilot coding and feedback from raters who were trained to use the rubrics. We then engaged in a generalizability study to understand how the rubrics were functioning. Based on the results of the initial generalizability study, we removed several rubrics due to a lack of variation in score distribution and we made revisions to several other rubrics to reduce rater variance or to improve the score distribution. We then engaged in a second generalizability study with the revised set of rubrics. The following four practices have rubrics that remained the same through both generalizability studies, and, hence, are the focus of this analysis: positioning students as competent; social coaching; supporting connection and engagement between student context and the math learning environment; and supporting a nurturing environment by proactively building relationships and productive classroom culture.

Four Focal Practices

Based on initial evidence of validity from the first generalizability study, the rubrics that correspond with the four focal practices had scores that demonstrated variability across lessons, and variability was mostly attributed to the lesson and not to differences across raters.

Positioning students as competent. Positioning students as competent is about teachers explicitly and publicly valuing, identifying, and acknowledging the brilliance of their students and framing their actions and statements as intellectually valuable (Bartell, 2011). Note, this is not "appointing" or "giving" students competence. Whether or not a teacher recognizes it, all students already have the capability and know-how to do important and brilliant things. The rubric that attends to this practice emphasizes the extent to which teachers specify what students do that is "productive" as well as the extent to which they provide rationales that support listening students as well as those being positioned in understanding why what was done was considered productive.

Social coaching. Coaching is one way teachers can support students in negotiating productive ways of participating and meeting expectations without decreasing the rigor of the task at-hand. Specifically, social coaching is about teachers deliberately intervening, scaffolding, or providing additional supports to help as students engage with one another (e.g., as they work in cooperative groups or present their thinking to one another). The rubric that attends to this practice focuses on the extent to which the teacher provides concrete suggestions in support of social participation. In addition, this rubric attends to how often the teacher provides rationales for their suggestions.

Supporting connection and engagement between student context and the mathematics learning environment. This practice is about connecting students' lives to discussions and interactions that take place in mathematics classrooms by making the most of connections between the mathematics discussed in class and the everyday lives of students. In particular, teachers may attend to aspects of students' lives and incorporate them into the curriculum (Ladson-Billings, 1995; Banks & McGee, 2001; Gay, 2002), or they may provide learning opportunities that make the mathematics problems discussed in class feel experientially real for students (Jackson, Garrison, Wilson, Gibbons, & Shahan, 2013).The rubric that attends to this practice focuses on the extent to which there are connections between mathematics and students' contexts, and whether the connections involve students through dialogue.

Supporting a nurturing environment by proactively building relationships and productive classroom culture. This practice is about establishing personal relationships and developing a sense of community in the classroom (Timmons-Brown & Warner, 2016). This practice often involves the teacher building rapport with students and reinforcing "classroom values." The rubric that attends to this practice emphasizes the extent to which the teacher connects with students in ways that are substantial or that are reciprocated. It also attends to the extent to which the teacher highlights or reiterates classroom values.

Method

Sample

This research project draws on extant classroom video data from two prior research projects, the Responsive Classroom Efficacy Study (RCES) study (Rimm-Kaufman et al., 2014) and the Middle-School Mathematics and the Institutional Setting of Teaching (MIST) study (Cobb et al., 2018). The RCES lessons included in this study were collected in upper elementary classrooms (grades 3-5) during the 2008–09, 2009-10, and 2010-11 school years, and the MIST lessons were collected in middle school classrooms (grades 6-8) during the 2009-10 and 2010-11 school years. The districts and schools within each district varied in their student demographics (see Table 1).

Table 1: District Student Demographics (Rounded)

District	Number of Students	% White	% Black	% Hispanic	% Free/ reduced price lunch
RCES	175,000	44	10	21	20
MIST-A	35,000	30	40	15	65
MIST-B	80,000	15	25	60	70
MIST-C	160,000	15	30	65	85
MIST-D	95,000	55	35	5	55

We hired and extensively trained 5 raters to use the EAR-MI rubrics. At the conclusion of several months of training, rater reliability was assessed with multiple measures including percent exact agreement with an expert score across the last five lessons rated, as well as across 21 lessons scored as a part of training. We consider the five most recently scored lessons because we expected agreement between the raters and expert to improve over time due to ongoing learning. Table 2 shows that exact agreement based on the last five relative to overall was higher for Positioning and Proactive, was similar for Social Coaching, and was slightly lower for Context.

Examining raters' scores relative to the expert scores and relative to each other across the 21 lessons included in training, we calculated Cohen's kappa, Fleiss's kappa, and Krippendorff's alpha (Table 2). Rater consistency with the expert tended to be higher than relative to one another. Generally speaking, kappa and alpha statistics above .20 indicate fair agreement and above .40 indicate moderate agreement (Klein, 2018). Agreement rates with the expert were generally fair or moderate. The Intraclass Correlation Coefficients (ICC) allow us to measure consistency in raters' scores relative to each other. Higher ICCs indicate that scores are trending in similar directions. The ICCs observed in our training data were all above .50, except social coaching which was .40.

For this analysis, we drew on data coded as part of two consecutive generalizability studies, and selected the four focal practices because the related rubrics remained the same and produced reliable results across the two generalizability studies. We analyzed a sample of 141 lessons, representing 65 teachers. This resulted in 83 upper elementary lessons, and 58 middle school lessons. For the purpose of this analysis, a set of scores was generated for each of these lessons through one of two different methods: 1) Expert scores, or 2) Averaging scores across raters. The expert scores resulted from lessons used for bi-weekly drift checks, and the other lessons were scored by three or more raters, and those scores were averaged to create a unique score for each rubric for each lesson. Rubrics include 5 discrete score points. By taking the average across raters, resulting scores could take on any value between 0 and 4.

Table 2: Measures of Rater Agreement

	Positioning	Social Coaching	Context	Proactive
Training				
% Exact Agreement (Last 5 Videos)	83.3%	53.3%	63.3%	73.3%
% Exact Agreement (Overall)	71.4%	53.2%	66.7%	55.6%
Relative to the Expert				
Cohen's Kappa	.53	.29	.53	.37
Fleiss's Kappa	.53	.29	.52	.35
Krippendorff's Alpha	.53	.29	.52	.35
Relative to Other Raters				
Cohen's Kappa	.39	.20	.41	.28
Fleiss's Kappa	.39	.19	.41	.28
Krippendorff's Alpha	.39	.19	.41	.28
Intraclass Correlation	.54	.40	.71	.51

Analysis

This study reports on a descriptive analysis of scores assigned by raters, based on four rubrics designed to assess the four focal practices: positioning students as competent (POSITIONING), social coaching (SOCIAL COACHING), supporting connection and engagement between student context and the mathematics learning environment (CONTEXT), and supporting a nurturing environment by proactively building relationships and productive classroom culture (PROACTIVE). With the unique set of scores for each lesson, we examined the score distributions for the four practices as well as compared score distributions for the middle and elementary school sub-samples. We utilized two-sample t-tests to determine whether perceived differences between the elementary and middle school sample means were statistically significant. A limitation of this analysis is that it does not take into account the nested nature of

the lessons within teachers or the order of the lessons for teachers. Future analyses will investigate the influence of these factors on the results included in this report.

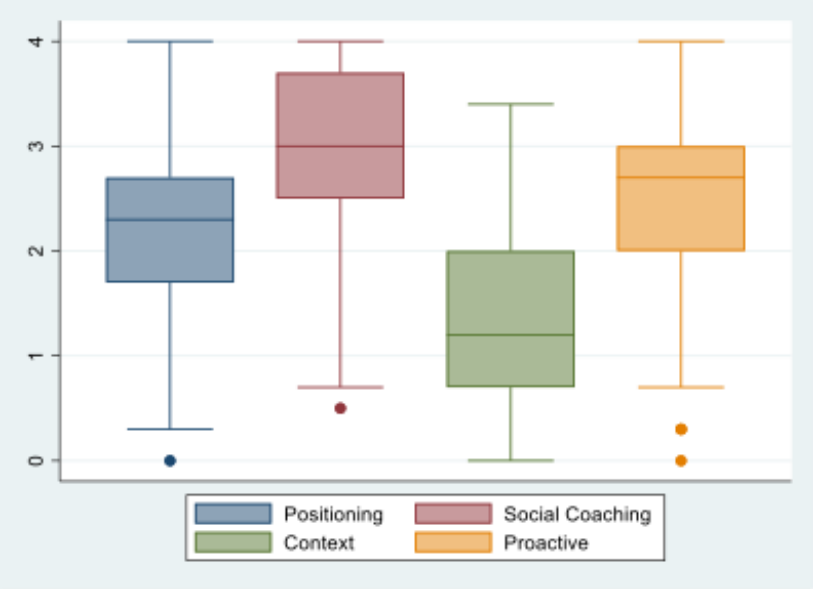

Figure 1: Box Plots Demonstrating Score Distributions for Four Focal Practices

Table 3: Descriptive Statistics for Scores Related to Four Focal Practices

	Overall (n=141) Mean (SD)	Middle (n=58) Mean (SD)	Elementary (n=83) Mean (SD)	Grade band T-test P-value
Positioning	2.20 (0.71)	2.05 (0.61)	2.31 (0.75)	p=.033*
Social Coaching	3.01 (0.75)	3.02 (0.80)	3.00 (0.72)	p>.05
Context	1.30 (0.87)	1.38 (0.92)	1.24 (0.84)	p>.05
Proactive	2.46 (0.67)	2.60 (0.43)	2.36 (0.79)	p=.035*

Note: * denotes statistically significant at the p<.05 level.

Results

Overall, the practices vary significantly in their score distributions (see Figure 1). Of the four focal practices, social coaching had the highest mean (m=3.01, see Table 3), and the score distribution is shifted considerably higher than for the other practices (see Figure 1). A mean of about 3 for social coaching can be interpreted as lessons in which the teacher provided concrete suggestions for social participation with occasional rationales for following those suggestions. The practice with the second-highest mean (m=2.46) was proactive (i.e., supporting a nurturing environment by proactively building relationships and productive classroom culture), which is between levels 2 and 3. For a level 2, the teacher made just one substantial attempt to connect with students (e.g., the teacher sharing personal information about their life). For a level 3, the teacher made more than one substantial attempt to connect with students. The practice with the next highest mean was positioning students as competent (m=2.20), which corresponds

between levels 2 and 3. At a level 2, we see lessons where the teacher positions at least one student as competent by specifying what the student did that was productive, but does not provide a rationale for why it was productive (e.g., "Great strategy"). At a level 3, the teacher provides rationales that may not be clear or that may not be focused on disciplinary practices of mathematics like generalizing, justifying, and making connections among multiple representations (e.g., "Using the lines on your paper is a great strategy because your work will be neat and structured and it will be easier to find your answer"). The focal practice with the smallest mean (m=1.30), but also the largest standard deviation (SD=0.87) was context (i.e., supporting connection and engagement between student context and the mathematics learning environment). At a level 1, either the connections to students' contexts are superficially related to the math task or the students do not participate meaningfully in the discussion of the context (e.g., with a math problem about the perimeter of a lake, the teacher might say, "Raise your hand if you have seen a lake before"). At levels 2 and above, the connections to students' contexts are substantially related to the math task, with different levels of student participation for levels 2-4. At a level 2, students participate using brief or one-word responses, and at a level 4, multiple students participate in developing a shared understanding of the connections to the context.

Dividing the sample into middle school and elementary school lessons revealed additional nuance with respect to some of the practices (see Table 3). First, there were no statistically significant differences between the elementary and middle school lesson sample means for context and social coaching. There were statistically significant differences for positioning and proactive, and they were in opposite directions. In particular, middle school lessons received significantly lower scores with respect to positioning students as competent ($p<.05$). A middle school mean at a level 2 indicates that, on average, middle school teachers did not provide rationales for their statements that positioned students as competent, whereas a mean of 2.31 for elementary lessons suggests that more elementary teachers provided some sort of rationale in their lessons. The significant difference between means was in the opposite direction for proactive. In other words, middle school lessons received higher scores with respect to supporting a nurturing environment by proactively building relationships and productive classroom culture. This means that there were more teachers in middle school lessons (when compared with elementary school lessons) who made more than one substantial attempt to connect with students. These grade band differences are interesting and warrant further investigation. In the discussion we offer several possible interpretations as well as implications for researchers and teacher educators.

Discussion

Our analysis of four practices that support equity and access in mathematics lessons highlighted interesting differences between the average middle school lesson and the average elementary school lesson with respect to positioning students as competent and proactively building relationships. On average, in elementary lessons compared with middle school lessons, more teachers tended to provide rationales when specifying what the students did that was productive while positioning them as competent. The rubric attends specifically to the explicit ways that teachers position students. With this in mind, it could be that middle school teachers position students as competent in ways that are mostly implicit and thus would not be documented as outlined in the rubric (e.g., some teachers position students by calling them up to the board and asking them to demonstrate their mathematical strategies in the front of the class). It is important to note, that we are not saying that implicit positioning cannot be useful. However, one theme that we have found across the practices (and especially with the practice of

positioning) is that the most supportive implementations of the practices usually reveal the often invisible "rules of the game" being played in mathematics classrooms that may not be apparent to students, particularly students who historically have been minoritized and marginalized in these contexts. In other words, what we have found is that the more transparent and explicit teachers can be in supporting their students the better. These transparent and explicit moves could support students in more directly accessing what is going on and what they are being asked to do, which may empower them in finding their *own* individual ways of "doing math".

On the other hand, in our sample, the average middle school lesson was rated significantly higher than the average elementary school lesson with respect to proactively building relationships. It could be that teachers in elementary classrooms connect with students in more superficial ways. Alternatively, we note that the distinctions within this rubric attend to the extent to which there are reciprocal personal connections and bonds being built between teachers and students while working on mathematics problems. Knowing that teachers of elementary-aged students tend to teach all subject areas and are usually the instructor for their students throughout the whole school day, it is possible that elementary teachers compartmentalize and make these types of substantial connections at other times during the school day (e.g., some teachers facilitate discussions about their own lives and inquire about their students' lives during "Calendar Time" or "Circle Time on the Carpet"). However, we are finding that these rich interpersonal connections are particularly important when teachers and students are working on mathematics as it is one way to support students in "showing up" completely and as their whole selves. These interactions also support students in viewing their teachers as approachable, which we have seen improve student participation both in terms of who participates and how they participate. In general, we find that these reciprocal interpersonal connections support the development of a space in which students, particularly those whose voices and natural ways of being are typically pushed to the margins, are likely to feel seen and to be comfortable being their authentic selves. These interactions also support relationships that help students feel safe and secure to take the necessary risks involved in "doing math" (e.g., knowing that disagreeing with shared ideas is common and can be non-threatening or knowing that they are free to make rough draft or not fully formed conjectures while discussing mathematics in class).

As we work to support mathematics teachers to go beyond high-quality mathematics instruction and specifically attend to equity and access in mathematics classrooms, we need more guidance about concrete practices that teachers can engage in (Grossman et al., 2013). This analysis is part of a larger effort to both specify those practices and develop research tools that can be used to assess teachers' progress as they work to provide instruction that aims towards equity in their mathematics classrooms. At this stage of the validation process, the attention in the EAR-MI rubrics is on whether the teacher engages in particular practices and not with which of the students the practices are enacted. Once we have established meaningful differences between scores on the rubrics, a possible extension would be to combine the EAR-MI rubrics with a participation-focused approach like the EQUIP (Reinholz & Shah, 2018) to attend to whether there are patterns in with whom the teacher enacts particular practices. By attending to individual students in the classroom, we could more intentionally address aspects of the critical axis, specifically highlighting how individual students experience specific aspects of instruction in a mathematics classroom.

The field of mathematics education needs additional research specifically focused on tools for researchers and practitioners that attend to the extent to which mathematics learning environments support historically marginalized groups of students. With the evidence that

teachers are not consistently enacting practices that support historically marginalized students, it is clear that teachers would benefit from learning more about these practices and how to enact them in mathematics classrooms. Teacher educators and professional development providers can use these practices and related rubrics to help teachers understand the practices and important distinctions in how they get enacted. By discussing important principles that guide work with students (e.g., the importance of building relationships with students), and pairing those with specific practices described by the rubrics, teachers can begin to envision how to enact those principles with students (e.g., see Pruitt-Britton et al., 2021).

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant #1908481. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Banilower, E. R., Smith, P. S., Malzahn, K. A., Plumley, C. L., Gordon, E. M., & Hayes, M. L. (2018). *Report of the 2018 NSSME+.* Chapel Hill, NC: Horizon Research, Inc.

Banks, J. A., & McGee, C. A. (2001). Families and teachers working together for school improvement. *Multicultural education: Issues and perspectives, 4*, 402-420.

Bartell, T. (2011). Caring, race, culture, and power: A research synthesis toward supporting mathematics teachers in caring with awareness. *Journal of Urban Mathematics Education, 4*(1), 50-74.

Gay, G. (2002). Preparing for culturally responsive teaching. *Journal of Teacher Education, 53*(2), 106-116.

Goodwin, C. (1994), Professional Vision. *American Anthropologist, 96*, 606-633.

Gutiérrez, R. (2012). Context matters: How should we conceptualize equity in mathematics education? In *Equity in discourse for mathematics education* (pp. 17-33). Springer, Dordrecht.

Gutstein, E., & Peterson, B. (Eds.). (2005). *Rethinking mathematics: Teaching social justice by the numbers*. Rethinking Schools.

Jackson, K., Garrison, A., Wilson, J., Gibbons, L., & Shahan, E. (2013). Exploring relationships between setting up complex tasks and opportunities to learn in concluding whole-class discussions in middle-grades mathematics instruction *Journal for Research in Mathematics Education, 44*(4), 646-682.

Klein, D. (2018). Implementing a general framework for assessing inter-rater agreement in Stata. *The Stata Journal: Promoting Communications on Statistics and Stata.* https://doi.org/10.1177/1536867X1801800408

Ladson-Billings, G. (1995). But that's just good teaching! The case for culturally relevant pedagogy. *Theory into Practice, 34*(3), 159-166.

Lampert, M., Beasley, H., Ghousseini, H., Kazemi, E., & Franke, M. (2010). Using designed instructional activities to enable novices to manage ambitious mathematics teaching. In *Instructional explanations in the disciplines* (pp. 129-141). Springer, Boston, MA.

Lampert, M., & Graziani, F. (2009). Instructional activities as a tool for teachers' and teacher educators' learning. *The Elementary School Journal*, *109*(5), 491-509.

Little, J. (2003). Inside teacher community: Representations of classroom practice. *Teachers college record*, *105*(6),

913-945.

Nasir, N. I. S., & Cobb, P. (2002). Diversity, equity, and mathematical learning. *Mathematical thinking and learning*, *4*(2-3), 91-102.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics.* Reston, VA: Author.

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common core state standards for mathematics*. Washington, DC: Authors. Retrieved from http://www.corestandards.org/assets/CCSSI_Math Standards.pdf

Pruitt-Britton, T., Wilhelm, A. G., & Wilson, J. (2022). Nurturing students through social interactions. *Phi Delta Kappan*, *103*(5), 18-23.

Reinholz, D., & Shah, N. (2018). Equity analytics: A methodological approach for quantifying participation patterns in mathematics classroom discourse. *Journal for Research in Mathematics Education, 49*(2), 140-177.

Rimm-Kaufman, S. E., Larsen, R. A. A., Baroody, A. E., Curby, T. W., Ko, M., Thomas, J. B., Merritt, E. G., Abry, T., & DeCoster, J. (2014). Efficacy of the Responsive Classroom Approach: Results from a 3-Year, Longitudinal Randomized Controlled Trial. *American Educational Research Journal, 51*(3), 567-603.

Timmons-Brown, S., & Warner, C. (2016). Using a Conference Workshop Setting to Engage Mathematics Teachers in Culturally Relevant Pedagogy. *Journal of Urban Mathematics Education*, *9*(1), 19-47.

Wilson, J. (Accepted). Initial steps in developing classroom observation rubrics designed around instructional practices that support equity and access in mathematics learning environments with potential for "success". *Teachers College Record: Special Issue*.

Wilson, J., Nazemi, M., Jackson, K., & Wilhelm, A. G. (2019). Investigating Teaching in Conceptually-Oriented Mathematics Classrooms Characterized by African American Student Success. *Journal for Research in Mathematics Education, 50*(4). 362-400. https://doi.org/10.5951/jresematheduc.50.4.0362