Title: The Contribution of Mathematics Instructional Quality and Class Size to Student Achievement for Third Grade Students from Low Income Families

Author(s): Eileen G. Merritt, Sara E. Rimm-Kaufman, Robert Q. Berry, Temple A. Walkowiak, and Ross A. A. Larsen

Authors 1, 2, 3, and 5: University of Virginia, Curry School of Education and Center for the Advanced Study of Teaching and Learning (CASTL)

Author 4: North Carolina State University at Raleigh

Background / Context:

Elementary students from low-income families often achieve well below their middle and high-income peers in mathematics despite efforts in the past two decades to reform mathematics education (NCES, 2009; Reardon, in press). The achievement gap between children from low-income families and children from higher income families is present when children enter school and does not appear to narrow appreciably throughout schooling (Duncan & Magnuson, in press). Further, the gap has steadily increased over the past 25 years (Reardon, in press), raising the prominence of the issue in school-based efforts to improve achievement.

Improving the quality of teaching has been raised as a promising focal point in efforts to improve mathematics achievement for all children, particularly children from low-income families (Hamre & Pianta, 2005; Jordan, Kaplan, Ola'h, and Locuniak, 2006; Nye, Konstapoulos & Hedges, 2004). Other school-based factors have been implicated as predictors of mathematics achievement in children from low income families as well. These include smaller class size (e.g. Krueger, 1999; Rockoff, 2003) and higher ability level of peers (Hanushek, Kain, Markman & Rivkin, 2003; Rockoff, 2003). However, many studies examining school-based factors on student achievement in mathematics have relied on value added models (e.g. Nye et al., 2004; Rockoff, 2003; Hanushek et al., 2003), requiring attention to what occurs *inside* of the classroom in order to guide future educational interventions.

Classroom observational measures can provide information about high quality student-teacher interactions, allowing researchers to consider the impacts of these practices on student outcomes. Such measures can take a *process-oriented approach* that considers the nature of interactions between teachers and students, such as the sensitivity of teachers' interactions with students, teachers' effective management of the classroom, and the depth of instruction and quality of feedback given to students—all processes that have been linked to achievement gains (Pianta, Belsky, Houts & Morrison, 2007; Pianta & Hamre, 2009; Ponitz, Rimm-Kaufman, Brock & Nathanson, 2009). Alternatively, observational measures can take a *domain-specific approach* that considers teachers' practices that support information processing in specific subject-areas, such as mathematical problem solving. Seidel and Shavelson conducted a recent meta-analysis that examined the effects of teaching on student learning, considering both domain-specific and process-oriented approaches to teaching in K-12 classrooms. They found that domain-specific processes had larger effects (d = .41) on cognitive outcomes in elementary classrooms than all other factors (Seidel & Shavelson, 2007). The present study combines process-oriented and domain specific approaches, using a newly developed measure: the Mathematics Scan (M-Scan) Measure of Mathematics Instructional Quality (Berry, Rimm-Kaufman, Ottmar, Walkowiak & Merritt, 2011).

Research Questions:

Three primary research questions guided our analyses: 1) What is the contribution of mathematics instructional quality to achievement for low-income students? We hypothesized that mathematics instructional quality is a strong predictor of mathematics instructional quality for low-income students, even after controlling for prior achievement, class-size, peer ability level, teachers' experience, and teachers' content knowledge, 2) What is the relative contribution of classroom ability level beyond classroom quality and class size in predicting achievement for low-income third grade students? We hypothesized that classroom ability level would contribute to achievement above and beyond mathematical instructional quality and other classroom factors, and 3) Does class size make a difference above and beyond instructional quality and classroom ability level in predicting achievement for low-income third grade students? Our hypothesis was that class size was significantly related to achievement above and beyond mathematics instructional quality and classroom ability level.

Setting:

Students and teachers from eleven schools in a large, suburban mid-Atlantic school district were studied. The schools were part of a larger randomized control efficacy study of a socio-emotional learning intervention, the *Responsive Classroom® (RC)* Approach *(*Northeast Foundation for Children, 2010). Data for the present study were collected from third grade students in 11 schools. These schools included students with 3 - 76 percent free and reduced lunch (*M* = .26), and 5 – 56 percent who were English-Language Learners (*M* = .25).

Participants:

All third grade children (n = 205) with designated free or reduced lunch status from 11 schools were included. Students were 105 male; 108 Hispanic, 19 Caucasian, 43 African American, and 27 Asian American, and 169 English Language Learners. Also, 30 students received special education services, including 17 who were learning disabled and 9 with speech or language disorders.

The students came from thirty-six different mathematics classrooms. Thirty five of the teachers reported demographic information, including 32 Caucasian, 3 African American, 1 Asian and 1 Native American. Thirty-two teachers were fully licensed and 3 were provisionally licensed. Teachers had a mean of 11 years experience (median = 9 years, range 1 – 35 years).

Intervention / Program / Practice:

Data for the present study were collected from teachers and students at control group schools. These schools were studied to examine the naturally occurring variation of teaching quality and other classroom factors in schools without systematic intervention.

Research Design:

The present study examines data from control schools participating in a three-year longitudinal cluster randomized control trial of the *RC* approach. Twenty-four study schools were selected from the collaborating district because of their interest in training in the *RC* approach. Roughly half of selected schools were assigned randomly (with stratification for ethnicity and free/reduced lunch composition) to intervention and control conditions. Control group schools were using "business as usual" approaches to classroom management and school-wide social and emotional learning.

Data Collection and Analysis:

Baseline measures of student achievement were conducted in the spring of 2008. Teachers were surveyed using online questionnaires in the spring of 2009, and reported on *demographic* *information* such as years' experience and level of certification. Part of the online survey included an assessment of *teachers' mathematics knowledge for teaching* (MKT; Hill, Shilling, & Ball, 2004). Teachers received $100 for completion of the questionnaire and MKT assessment and for permitting classroom observations. Classroom observations were conducted and videotaped by research assistants in all third grade classrooms at three time points during the 2008-2009 school year. Observations of full math lessons (approximately one hour in duration) occurred once in the fall (September to November), once in winter (December to February), and once in spring (March to May). *Class size* was reported by teachers during each observation. Finally, achievement tests in mathematics were conducted in the spring of third grade. Student *demographic information* was collected concurrently with achievement tests by school administrators. See Table 1 for a complete list of measures and their sources.

Classroom observations were watched and scored by trained research assistants using the M-Scan (Berry et al., 2010). The M-Scan defines 8 aspects of *mathematics instructional quality* including: *structure of the lesson; multiple representations; students' use of mathematical tools; cognitive depth; mathematical discourse community; explanation and justification; problem solving; and connections and applications.* These dimensions are tied to National Council of Teachers of Mathematics (NCTM) Standards (2000), and capture some domain-specific processes that occur in mathematics classrooms such as problem-solving and the use of mathematical representations. To develop the M-Scan, the RCES research team relied on prior work by Borko and others (Borko, Stecher, Alonzo, & McClam, 2005) in the SCOOP measure, and Pianta and others in the CLASS measure (Pianta, LaParo, & Hamre, 2007). All coders attained reliability of 80% within-one of master codes for each dimension prior to beginning coding; monthly drift tests helped coders maintain high reliability throughout the study. Internal consistency for the 8 dimensions was strong, $\alpha = .93$. Work establishing the validity of this measure is currently under review (Walkowiak, Berry, Meyer, Rimm-Kaufman & Ottmar, under review).

Assumptions of normality, linearity, and homogeneity of variance were examined and met. Two outliers were identified and were removed. Missing data were handled using Full Information Maximum Likelihood approach in MPlus (Muthén, B., & Muthén, L. K., 2010). Analyses were conducted in 4 steps. Step one: Descriptive statistics and bivariate correlations were used to examine missing data, model assumptions, and associations among independent and dependent variables. Step two: Unconditional models were estimated using multilevel models in *MPlus* software to account for the nesting of children within classrooms (Muthén & Muthén, 2008). Two-level models were selected in accordance with our research questions focused on variability in teacher quality. The intraclass correlation coefficient (ICC) was 10.4% for third grade achievement. Step three: Models were built incrementally, adding a level 1 control variable (Stanford 10 second grade mathematics score), level 2 control variables (years of *teaching experience*, *mathematics knowledge for teaching*), and level 2 predictors (*classroom ability level*, *class size*, and *mathematics instructional quality*). The level 1 and 2 variables were centered at the grand mean to control for their influence in the model and interpret results more easily (Enders & Tofighi, 2007). Step 4: Finally, to quantify the local effect size for mathematics instructional quality and class size, each variable of interest was added as the last predictor in the model, and the proportion of variance reduction (PVR) was calculated by comparing the final model to a model with all of the variables except for the variable of interest.

Results:

Descriptive statistics and correlations for all variables can be found in Table 2. The results of the hierarchical linear model can be found in Table 3. The final model explains 43.9% of the total variance in student achievement, including 91.56% of the classroom level variance.

RQ 1: We found that mathematical instructional quality was significantly related to third grade achievement for students from low-income families. For every one point higher a teacher scored on the M-Scan, students scored approximately 12 points higher on a third grade achievement test. Mathematical instructional quality explains 8% of the classroom level variance and .4% of the total variance in student achievement after controlling for student prior achievement, teachers' content knowledge, class size, classroom ability level, and teachers' experience level.

RQ 2: Results showed the classroom ability level was not a significant predictor of student achievement with all of the other variables in the model. This suggests that when students are offered the same level of instructional quality, teacher content knowledge, teacher experience, and have similar class sizes, the ability level of peers is not significant.

RQ 3 Class size was a significant predictor of achievement for students from low-income families. For every 3 fewer students in a classroom, students scored 11 points higher on the third grade achievement test. Class size explained 28% of the classroom level variance and 3% of the total variance in achievement above and beyond the other variables in the model.

Conclusions:

Results support prior work suggesting that high quality teachers and lower class sizes are both important for third grade students who come from low-income families. Specifically, teaching practices that are aligned with NCTM standards are related to higher achievement for these students. This study extends prior work from value-added models by examining the effects *of mathematics instructional quality* in addition to class size, teacher experience, mathematics content knowledge, and classroom ability level in predicting mathematics achievement.

Two limitations are worth noting. First, participants in our study came from a large, well-funded suburban school district. In this district, administrators are able to adjust class sizes and personnel to classrooms where students need more help. These results may not generalize to samples in large urban districts that are underfunded. Second, the M-Scan measure is a recently developed measure. Initial findings support its validity and usefulness (Berry, Rimm-Kaufman, Ottmar & Merritt, 2011; Walkowiak, Berry, Rimm-Kaufman, Meyer & Ottmar, under review). However, results from analyses with any new measure should be interpreted cautiously.

Although the design does not permit causal inferences, findings suggest the importance of *mathematics instructional quality* and smaller class size among students from families with low income. Interventions that support teacher improvement in standards-based mathematics may hold promise in efforts to reduce the achievement gap. For example, professional development for teachers that promotes high quality mathematical discourse or multiple approaches to problem solving may be beneficial. Future work on teacher quality should use observational measures that capture the nature and quality of domain-specific processes in classrooms, and consider whether teachers are better able to offer higher *mathematics instructional quality* when they have smaller classes.

Appendices

Appendix A. References

Berry, III, R. Q., Rimm-Kaufman, S. E., Ottmar, E. M., Walkowiak, T. A., & Merritt, E. (2010). *The Mathematics Scan (M-Scan): A Measure of Mathematics Instructional Quality*. Unpublished measure, University of Virginia.

Berry, R.Q., Rimm-K Kaufman, S., Ottmar, E., & Merritt, E. (2011). The M-Scan Measure of Mathematics Instructional Quality. Interactive paper session presentation at the National Council of Teachers of Mathematics (NCTM) Conference, Indianapolis, Indiana.

Borko, H., Stecher, B. M., & Alonzo, A. C, & McClam, S. (2005). Artifact packages for characterizing classroom practice: A pilot study. *Educational Assessment, 10*(2), 73-104.

Duncan, G., & Magnuson, K. (in press). In R.M. Murnane & G. Duncan (Eds.), *Whither Opportunity? Rising Inequality and the Uncertain Life Chances of Low-Income Children*. Washington, D.C.; Brookings Institution.

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12*(2), 121-138.

Hamre, B. K. & Pianta, R. C. (2005). Can instructional and emotional support in the first-grade classroom make a difference for children at risk of school failure? *Child Development, 76*(5), 949-967.

Hanushek, E.A., Kain, J. F., Markman, J. M., & Rivkin, S. G. (2003). Does peer ability affect student achievement? *Journal of Applied Econometrics, 18*, 527 – 544.

Harcourt Educational Meaurement. (2008). Stanford Achievement Test – Tenth Edition, Second Grade Mathematics.

Hill, H.C., Schilling, S.G., & Ball, D.L. (2004) Developing measures of teachers' mathematics knowledge for teaching. *Elementary School Journal (105)*, 11-30.

Jordan, N.C., Kaplan, D., Ola'h, L.N., & Locuniak, M.N. Number sense growth in kindergarten: a longitudinal investigation of children at risk for mathematics difficulties. *Child Development 77*(1), 153 – 175.

Krueger, A. B. (1999). Experimental estimates of education production. *The Quarterly Journal of Economics, 114*(2), 497-532.

Muthén, L.K. and Muthén, B.O. (1998-2010*). MPlus User's Guide,* sixth edition. Los Angeles, CA: Muthén & Muthén

Muthén, B., & Muthén, L. K. (2010). MPlus (Version 6.0). Los Angeles, CA: Muthén & Muthén.

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis, 26*(3), 237.

National Center for Education Statistics (2009). The Nation's Report Card: Mathematics 2009 (NCES 2010–451). Institute of Education Sciences, U.S. Department of Education, Washington, D.C.

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

Northeast Foundation for Children. (2010). Retrieved from http://www.responsiveclassroom.org/

Pianta, R. C., Belsky, J., Houts, R., & Morrison, F. (2007). TEACHING: Opportunities to learn in America's elementary classrooms. *Science*, *315*(5820), 1795.

Pianta, R.C. & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: standardized observation can leverage capacity. *Educational Researcher*, *38*(2), 109 – 119.

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2007). Classroom assessment scoring system, manual, K–3. Baltimore: Paul H. Brookes Publishing Co., Inc.

Ponitz, C.C., Rimm-Kaufman, S.E., Brock, L. L. & Nathanson, L. (2009). Early adjustment, gender differences, and classroom organizational climate in first grade. *Elementary School Journal 110*(2), 142 – 162.

Reardon, S. F. (in press). The widening academic achievement gap between the rich and the poor: new evidence and possible explanations. In R.M. Murnane & G. Duncan (Eds.), Whither Opportunity? Rising Inequality and the Uncertain Life Chances of Low-Income Children. Washington, D.C.; Brookings Institution.

Rockoff, J.E. (2003). The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data.

Seidel, S., & Shavelson, R.J. (2007). Teaching effectiveness research in the past decade: the role of theory and research design in disentangling meta-analysis results. *Review of Educational Research, 77*(4) 44 – 499.

Sanders, W., & Rivers, J. (1996). *Cumulative and residual effects of teachers on future student academic achievement*. Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center.

Virginia Department of Education (2009). Spring 2009 Released test: Grade 3 mathematics.

Walkowiak, T.A., Berry, R.Q., Meyer, J.P., Rimm-Kaufman, S.E., & Ottmar, E.R. (under review). Introducing an observational measure of mathematics instructional quality: Evidence of validity and score reliability.

Wright, S.P., Horn, S.P., & Sanders, W.L. (1997). Teacher and classroom context effects on student achievement: implications for teacher evaluation. *Journal of Personnel Evaluation in Education, 11*(1), 57 – 67.

Appendix B. Tables and Figures

Table 1
Constructs and Measures

Construct	Measure
Prior Mathematics Achievement	Students took the Stanford Achievement Test- Tenth Edition (Stanford-10; Harcourt Educational Measurement, 2008) in the spring of second grade.
Mathematics Knowledge for Teaching	Teachers took the numbers and operations portion of the MKT (Hill, Shilling & Ball, 2004) in the spring of 2009. Scoring is based on national mean scores.
Classroom Ability Level	A classroom mean for the Stanford 10 test was computed to determine average ability level within each mathematics classroom.
Class Size	Class size data were collected during classroom observations throughout the third grade year.
Mathematics Instructional Quality	Research assistants rated videotaped complete math lessons at 3 points during the year using M-Scan (Berry, Rimm-Kaufman, Ottmar, Walkowiak & Merritt, 2010). Each item was coded on a 1-7 Likert scale, including low (1, 2), medium (3, 4, 5) and high (6, 7) quality instruction, and scores were averaged for each classroom.
Outcome: Third Grade Mathematics Achievement	Students took the Virginia Standards of Learning mathematics assessment in the spring of third grade (Virginia Department of Education, 2009).

Table 2
Summary of Correlations, Means and Standard Deviations for All Variables

Measure	1	2	3	4	5	6	7
N	203	196	196	205	205	205	198
Mean	559.06	.02	10.69	567.78	19.25	3.37	460.9
Standard Deviation	30.62	.75	8.37	18.94	4.00	.81	64.19
Minimum	494	-1.16	1	443.93	11	1.75	296
Maximum	662	1.19	35	624.50	31	4.75	600
1. Second grade achievement							
2. Mathematics knowledge for teaching	.04						
3. Years' experience teaching	.21***	-.27***					
4. Classroom ability level	.40***	.05	.29***				
5. Class size	.21***	.23***	.08	.46***			
6. Mathematics instructional quality	-.11	.60***	.04	-.15*	-.08		
7. Third grade mathematics achievement	.61***	.07	.19*	.19***	-.06	.09	

Table 3

Hierarchical Linear Model Results: Class Size and Mathematics Instructional Quality Predict Achievement for Third Grade Children from Low-Income Families

Parameters	Unconditional Model		Final model	
	Coef.	Standard Error	Coef.	Standard Error
	Fixed effects			
Level 1				
Intercept (γ_{00})	460.04***	4.58	458.91***	4.25
Stanford 10 (γ_{10})			1.38***	.15
Level 2				
MKT			.55	5.7
Classroom ability			.18	.24
Years experience			.45	.54
Mathematics instructional quality			11.84**	4.36
Class size			-3.66***	.87
	Random effects			
Intercept (τ_{00}) Level 2	429.26*	278.32	36.25	113.24
Intercept (σ^2) Level 1	3686.13***	409.32	2240.79***	262.42