Abstract Title Page
Not included in page count.

Title: Applying New Methods to the Measurement of Fidelity of Implementation: Examining the Critical Ingredients of the Responsive Classroom Approach in Relation to Mathematics Achievement

Author(s): Tashia D. S. Abry, Sara E. Rimm-Kaufman, Ross A. Larsen, & Alix J. Brewer; University of Virginia, Center for Advanced Study of Teaching and Learning

Background / Context:

Randomized controlled trials (*RC*T) conducted without sufficient attention to program implementation fall short of efforts in education science to understand the underlying mechanisms of interventions intended to improve student achievement. There has been a surge of interest on the assessment of implementation in order to better understand the processes underlying program efficacy. Fidelity of implementation (FOI) is the extent to which an innovation is implemented as intended (Dusenbury, Brannigan, Falco, & Hansen, 2003). Existing research describes a well-established link between FOI and program effects; programs with high FOI demonstrate effect sizes two to three times greater than programs with low FOI (Durlak & Dupre, 2008). However, when employed, FOI measures have been highly variable in their reliability, validity, and breadth (Durlak & Dupre, 2008; O'Donnell, 2008). Improving the quality of FOI measures is a crucial aspect of unpacking the "black box" of an intervention's causal mechanisms, including identification of its key ingredients (Century, Rudnick, & Freeman, 2010).

As the science of implementation matures, so are the approaches to assessing it. Nelson, Cordray, Hulleman, Darrow, and Sommer (2009) proposed a five-step "model-based" approach to the assessment of FOI: 1) *Specify the change model*. The change model refers to the constructs involved in the causal processes by which the intervention is hypothesized to impact outcomes. This model outline the contents of the black box and informs the components to be measured. 2) *Develop valid and reliable measures of model/intervention components (e.g., measures of FOI)*. Such measures may include implementer reports, observations, products, and logs. 3) *Employ the valid and reliable measures in data collection*. Ideally, indices are administered to both treatment and control groups, though adaptation for control participants may be required. 4) *Combine indices*. Indices may be combined in a variety of ways. Survey responses may be averaged across items to create a composite score (referred to here as the measure-based approach). As another example, multiple indicators of a single practice/activity may be combined to facilitate comparison of level of implementation of different program components (referred to here as the practice-based approach). 5) *Link the FOI measures to outcomes of interest*. Though not causal in nature, linking FOI indices to outcomes informs the extent to which variability in FOI relates to variability in outcomes, the interpretation of which will vary dependant on the index used.

In the context of an *RC*T evaluating the *Responsive Classroom®* (*RC*) approach, the present paper employs the five-step model proposed by Nelson and colleagues, in an effort to distill the critical ingredients of *RC* influential to students' mathematics achievement. *RC* is a social emotional learning (SEL) intervention designed to support students' socioemotional and academic skills (Northeast Foundation for Children, Inc. [NEFC], 2010), and combines teacher practices aimed at promoting high quality instructional strategies, students' self-regulation and executive function, and a positive classroom climate. The goals of the *RC* approach are well aligned with the process standards formulated by the National Council on Teachers of Mathematics (NCTM, 2000) and research has demonstrated the link between the types of practices advanced by *RC* and students' mathematics achievement (Blaire & Razza, 2007; Duncan, Dowsett, Claessens, Magnuson, Huston, Klebanov et al., 2007). In Figure 1, we offer a change model describing the potential impact of the *RC* approach on students' mathematics achievement. Pursuant to the goal of unpacking the mechanisms by which the use of *RC* practices influences students' mathematics achievement, we combine indices of FOI of the *RC* approach and compare both measure-based and practice-based composites in their prediction of student mathematics achievement.

Purpose / Objective / Research Question / Focus of Study:

Pending acceptance the presentation will describe *RC*ES in relation to the five step model. For brevity, we describe analyses pertaining to Steps 4 and 5, addressing two primary research questions:

1) Taking a traditional measure-based approach, we ask: Are observed and teacher reported measure-based composites of FOI of the *RC* approach predictive of students' fourth grade mathematics achievement?
2) Applying a contemporary practice-based approach, we ask: Are indicators of specific *RC* practices, compiled across observed and teacher reported FOI measures, predictive of students' fourth grade math achievement?

Setting:

Twenty-four schools from a single district in the mid-Atlantic were randomly assigned to a treatment or control group after stratifying on percentage of student eligibility for free/reduced lunch and minority student composition. Randomization yielded 11 schools assigned to the control group and 13 schools assigned to an experimental group. Assessment at randomization demonstrated that participating schools were demographically diverse in their representation of student free /reduced lunch status (Range = 2% - 72%, *M* = 26%) and minority student composition (Range 17% - 86%, *M* = 55%), and that treatment and control schools did not differ significantly on these variables. School size ranged from 289 to 986 students and percent of ELL students ranged from 5% to 75%, also with no significant differences between treatment and control schools.

Population / Participants / Subjects:

The teacher sample included 100 fourth grade teachers from the 24 schools, representing a 96% response rate across treatment and control groups. Table 1 provides teacher demographic characteristics according to group assignment. T-tests revealed no significant differences between the teacher groups on the tabled characteristics with the exception of race.

The student sample includes 2,266 fourth grade students from the 24 participating schools, including all students eligible for state testing. Table 2 displays student demographic characteristics by treatment and control. T-tests revealed that students in the treatment group were significantly more likely to have free/reduced lunch status and had lower mathematics scores at pre-test. Between the 2008-09 and 2009-10 school years, child attrition was 13% in the intervention group and 12% in the control group.

Intervention / Program / Practice:

RC is an approach to elementary teaching designed to create classrooms conducive to the optimal development of students' social and academic skills (NEFC, 2010). Informed by education research and theory, seven guiding principles emphasizing the importance of social development, understanding students as individuals, and school-family relationships provide the framework for the *RC* approach. Ten *RC* practices emanate from these the guiding principles and are described in Figure 2. See www.responsiveclassroom.org for more information.

Participating fourth grade teachers in the experimental group attended one-week training institutes during two consecutive summers in 2008 and 2009. The first session introduced the *RC* practices of Morning Meeting, Rule Creation, Logical Consequences, Interactive Modeling, and Positive Teacher Language, while the second session incorporated the remaining practices. In addition, treatment teachers received two to three coaching sessions with *RC* personnel throughout each of the 2008-09 and 2009-10 school years. Counterparts in the control group received no exposure to *RC* training or coaching support and continued "business as usual."

Research Design:

The present study examines data collected during the second year of a three-year longitudinal cluster randomized controlled trial, the *Responsive Classroom* Efficacy Study (*RC*ES). In the context of and *RC*T, our research questions address naturally occurring variability in our independent variables of interest (i.e., teachers' FOI to *RC* practices) and their relation to students' mathematics achievement.

Findings highlight the differences between traditional measure-based composites of FOI and contemporary practice-based composites, as well as the relation between teachers' use of specific *RC* practices and students' mathematics achievement, but do not permit causal inferences regarding these associations.

Data Collection and Analysis:

Data collected were from student achievement tests, classroom observations, teacher surveys, and district records. The Virginia Standards of Learning (SOL) test in mathematics administered in the spring of 2009 (in third grade) served as a baseline, with fourth grade scores collected the following year used as the dependent variable. Model controls were drawn from district records acquired during the 2010 school year.

Three measures of teachers' FOI of *RC* practices were developed and administered for the present study. The *Classroom Practices Teacher Survey* (Nathanson, Sawyer, & Rimm-Kaufman, 2007; [CPTS]) is a 46-item teacher-reported assessment of the use of *RC* practices (α = .86 and .93, baseline and post-test, respectively). CPTS items are worded without the use of *RC* vernacular enabling the measure to be administered to control teachers, and to minimize response bias of treatment teachers. Teachers are asked to reflect on the year and to report the extent to which each item was characteristic of their classroom on a one to five likert scale (see Figure 3 for examples of items). The CPTS was collected at baseline for all teachers in the spring of 2008, prior to any exposure to *RC* training for the treatment group. In the spring of 2010, these teachers completed the CPTS a second time. The post-test scores, averaged across items to create a single composite score, served as one measure-based indicator of teacher reported FOI of the *RC* approach.

A second teacher-reported measure of the use of *RC* practices, administered concurrently with the CPTS post-test, was the *Classroom Practices Frequency Survey* (Nathanson, Sawyer, & Rimm-Kaufman, 2007; [CPFS]). The CPFS (α = .89) is an 11-item survey in which teachers are asked to reflect on the first three months of the school year and to report on a one to eight likert scale the frequency with which they conducted each practice described (Figure 3). Similar to the CPTS, these scores, averaged across items, served as a second measure-based composite of teacher reported FOI.

The third implementation assessment collected was the *Classroom Practices Observation Measure* (Abry, Brewer, Nathanson, Sawyer, & Rimm-Kaufman, 2010; [CPOM]). The CPOM is an observational measure of teachers' use of *RC* practices (α = .89) rated on a three-point likert scale (Figure 3). Treatment and control teachers were observed for 60 minutes on five separate occasions throughout the 2009-10 school-year; three times during mathematics instruction (10 items) and twice in the morning during the first hour of school (with an additional 6 items). Intra-class correlations reflecting the inter-rater reliability of CPOM observers ranged from .74 to .88. Individual items were averaged across the five observations and then aggregated to yield a single score of observed FOI, serving as the third measure-based composite.

Pursuant to the goal of distilling the specific *RC* practices most influential to mathematics achievement, items assessing three *RC* practices, Morning Meeting (MM), Academic Choice (AC), and Interactive Modeling (IM) were selected from each of the above measures and combined using a meta-factor analytic process to create a corresponding practice-based composite. In the first step, the groups of items from a given measure corresponding to each of the aforementioned practices were factor analyzed to assess internal consistency and unidimensionality. The resulting factor scores were retained for each individual teacher such that each had three factor scores for each practice ($M = 0$, $SD = 1$) corresponding to items drawn from the three measures. In the second step, each set of factor scores for a given practice were factor analyzed to create a single factor score representing the teacher's implementation of the practice relative to the overall average. In the final step, confirmatory factor analyses were conducted in MPLUS (Muthen & Muthen, 1998-2009) to ensure satisfactory model fit for each of the three practice-based composites. This process accounted for the difference in scaling between the three measures as well as the disparity in number of items representing each practice. Furthermore, the use of factor loadings ensured items were weighted according to their statistical contribution to the underlying factor. Table 3 provides relevant details related to this process including number of items in each composite, internal consistency values, and fit indices. Descriptive statistics for the measure- and practice-based composites are reported in Table 4.

For each research question, two-level models in which students were nested in classrooms were analyzed using MPLUS. Models were specified using TYPE=COMPLEX TWOLEVEL to account for the third level of nesting within schools, and school level covariates including treatment assignment were treated as level two covariates.[1] All models were run as means-as-outcomes models with only the level one intercept for each outcome treated as random. With the exception of those for which zero was a meaningful value, all covariates were grand-mean centered to assist with interpretability.

Findings / Results:

Complete results for both measure- and practice-based composites are presented in Table 5. All models controlled for child gender, age, third grade math SOL scores, free/reduced lunch status, and ELL status at level 1, and treatment assignment, masters degree, years teaching experience, percent free/reduced lunch, school size, and percent ELL at level 2. Results indicated no significant relation between the measure-based composites and students' fourth grade mathematics achievement. R-squared values were .46 at the within level and .38 at the between level indicating the percent of variance in math achievement explained by the model.

Results of the practice-based model revealed that the Academic Choice composite was significantly predictive of students' mathematics achievement ($t = 2.41$, $p = .02$). A one standard deviation increase in teachers' AC composite scores was associated with a .24 standard deviation increase in students' math SOL score, equivalent to a 17 point increase. R-squared values for the practice-based model increased from the measure-based model to .46 at the between level and .52 at the within level.

Conclusions:

First, and most concretely, the findings suggest that giving students some choice in how they engage in their mathematics work appears to link to later mathematics achievement. Most likely, these aspects of choice offer students needed autonomy in the classroom and thus, meet a basic developmental need. Further, the opportunity to choose activities may support students' engagement in learning, which in turn, may foster motivation.

Second, the method proposed by Nelson et al (2009) offers an opportunity to uproot a default approach commonly used in the field. Researchers commonly assume that the high internal consistency of a measure implies that the composite derived will be maximally useful in subsequent analyses. Developing composite scores that cut across various measures may provide a useful way to operationalize new constructs that may have more meaning and greater predictive validity than the traditional approaches. As evidenced by the results reported here, the practice-based approach outperformed the measure-based approach in predicting and accounting for variance in students; mathematics achievement.

Finally, the approach presented here is congruent with recent goals of education science to establish the evidentiary basis for educational practices. The goal of RCTs is to test the impact of an intervention, but analysis of its components and the mechanisms underlying its effectiveness is essential towards the goal of developing increasingly effective interventions. (Schneider, Carnoy, Kilpatrick, Schmidt, & Shavelson, 2007).

Footnotes:
[1] All models were also run with children nested in schools, the level of randomization. The pattern of results was consistent with those presented here.

Appendices

Not included in page count.

Appendix A. References

Abry, T., Brewer, A., Nathanson, L., Sawyer, B., Rimm-Kaufman, S.E. (2010). Classroom Practices Observation Measure. University of Virginia, Unpublished measure.

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development, 78,* 647-663.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P. et al. (2007). School readiness and later achievement. *Developmental Psychology, 43*, 1428-1446.

Durlak, J. A. & Dupre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*, 327-350.

Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research, 18*(2), 237-256.

Muthén, L.K. and Muthén, B.O. (1998-2009). Mplus User’s Guide. Fifth Edition. Los Angeles, CA: Muthén & Muthén.

Nathanson, L., Sawyer, B., & Rimm-Kaufman, S.E. (2007). *Classroom Practices Teacher Survey*. Unpublished measure. University of Virginia.

Nathanson, L., Sawyer, B., & Rimm-Kaufman, S.E. (2007). *Classroom Practices Frequency Survey*. Unpublished measure. University of Virginia.

O’Donnel, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research*, *78*(1), 33-84.

Schneider, B., Carnoy, M., Kilpatrick, J., Schmidt, W. H., & Shavelson, R. J. (2007). *Estimating causal effects using experimental and observational designs*. American Education Research Association.

Appendix B. Tables and Figures
Not included in page count.

Figure 1
RC Change Model of Potential Impact on Mathematics Achievement

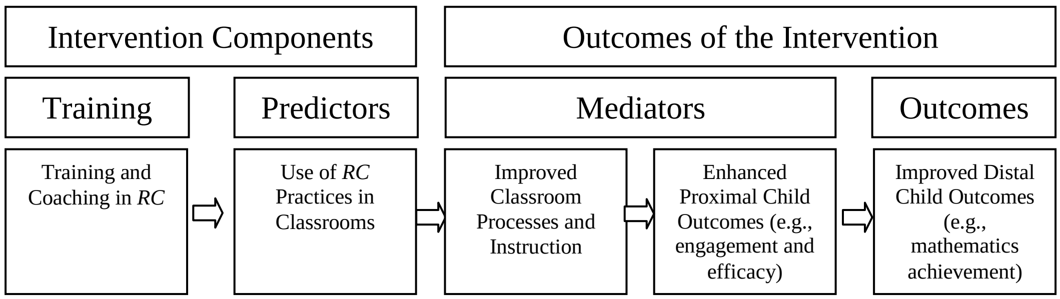

Table 1
Teacher Demographics

Demographic characteristic	Control (*n* = 45)			Treatment (*n* = 55)		
	%	*M*	*SD*	%	*M*	*SD*
Female	94			90		
Age		39	13		39	11
Caucasian	90			80		
Years teaching experience		12	11		10	7
Has masters degree	73			67		

Table 2
Child Demographics

Demographic characteristic	Control (*n* = 1,077)			Treatment (*n* = 1,189)		
	%	*M*	*SD*	%	*M*	*SD*
Female	50			53		
Age		10	0.40		10	0.40
Free/reduced lunch	28			34		
ELL status	38			40		
Baseline math SOL		498	69		479	75

Figure 2
Description of *RC* Practices

RC practice	Description
Morning Meeting	Daily meeting designed to promote relationships and positive climate consisting of a greeting, activity, sharing, and an interactive message
Rule Creation	Process by which teacher and students work together to distill a set of rules supportive of student generated social and academic goals
Interactive Modeling	Multi-step approach used to instruct children in expected classroom behaviors consisting of demonstration, observation, and practice
Positive Teacher Language	Communication used to promote student accountability and effort including reinforcing, reminding, and redirecting language
Logical Consequences	Relevant, respectful, and realistic consequences directly related to student misbehavior
Guided Discovery	Introduction of classroom materials promoting student autonomy, creativity, and responsibility
Academic Choice	Structured and monitored opportunities for students to choose, reflect on, and share work options based on individual interests
Classroom Organization	Classroom structures that encourage independence and productivity and highlight student contributions
Working with Families	Two-way communication in which families and teachers work together to identify individualized goals
Collaborative Problem Solving	Creative and effective strategies for resolving problems with and among students

Figure 3
Examples of CPTS, CPFS, and CPOM items

Measure	Scale	Example items
CPTS	5-point: Not at all characteristic - Extremely characteristic	Students greet each other by name during class meetings.
		When a rule is introduced, I ask students to model what following the rule looks like.
		I provide students a set of choices about what kind of work to do, how to do the work, or both.
CPFS	8-point: Almost never - More than once per day	I prepare a message on a chart/blackboard to which students are expected to respond.
		When a rule is introduced, I demonstrate to students how to correctly follow the rule or procedure.
		I help students plan their chosen work (e.g., provide examples of animals they can study, record their choices, and discuss their choices with them)
CPOM	3-point: Not at all characteristic - Very characteristic	The class has an established space for a morning meeting where all students are able to sit comfortably and have a clear view of classmates and relevant materials.
		Teacher provides explicit instruction (including teacher or student demonstration, student observations, and student practice) on how to do routine behaviors.
		Students make individualized choices related to an academic lesson or goal. Choices may be about the content or process of their academic work.

Table 3

Internal Consistency and Fit Indices for Fidelity Practice-based Composites

Practice Composite	Measure used	Number of items used	Alpha	Composite alpha	CFA fit indices for composite
MM	CPTS	12	0.97	0.94	CFI = .99; TLI = .97; RMSEA = .10; SRMR = .09
	CPFS	5	0.94		
	CPOM	5	0.92		
AC	CPTS	3	0.85	0.51	CFI = .92; TLI = .84; RMSEA = .09; SRMR = .06
	CPFS	3	0.89		
	CPOM	2	0.84		
IM	CPTS	5	0.75	0.53	CFI = .94; TLI = .89; RMSEA = .05; SRMR = .04
	CPFS	3	0.91		
	CPOM	0[1]	NA		

Note: MM = Morning Meeting; AC = Academic Choice; IM = Interactive Modeling

Footnote:

[1] The CPOM contained one item assessing teachers' observed use of Interactive Modeling, but was omitted to improve model fit.

Table 4
Correlations and Descriptive Statistics for Fidelity Measure-based and Practice-based Composites

Variable	1	2	3	4	5	6	7
Measure-based composites							
1. CPTS	-						
2. CPFS	0.83***	-					
3. CPOM	0.72***	0.59***	-				
Practice-based composites							
4. MM	0.88***	0.87***	0.77***	-			
5. AC	0.53***	0.55***	0.30***	0.27***	-		
6. IM	0.52***	0.57***	0.27***	0.30***	0.45***	-	
Outcome							
7. Math SOL scores	0.01	< .01	-0.01	-.08**	0.16***	0.03	-
M	3.60	4.76	1.61	-0.14	0.07	-0.07	512.15
SD	0.60	1.53	0.32	1.05	0.95	1.07	71.51
Min	2.26	1.45	1.02	-1.73	-2.48	-2.46	281
Max	4.74	7.55	2.30	1.11	2.39	1.91	600

** $p \leq .01$, *** $p \leq .001$

Note: CPTS = Classroom Practices Teacher Survey; CPFS = Classroom Practices Frequency Survey; CPOM = Classroom Practices Observation Measure; MM = Morning Meeting; AC = Academic Choice; IM = Interactive Modeling

Table 5

Results of Measure- and Practice-based Composites Predicting Fourth Grade Mathematics Achievement

	Coefficient	SE	Standardized coefficient
Measure-based model			
Fixed effects			
Intercept	519.98***	5.71	22.16
CPTS	10.5	6.76	0.27
CPFS	-2.14	2.45	-0.15
CPOM	8.12	11.06	0.12
Random effects	Variance	SE	Standardized coefficient
Intercept variance (u_{oj})	341.77***	87.99	0.62
Level 1 variance (r_{ij})	2240.96***	86.51	0.54
Practice-based model			
Fixed effects			
Intercept	513.73***	5.15	20.57
MM	-7.14	3.79	-0.3
AC	6.73*	2.8	0.24
IM	2.01	1.67	0.09
Random Effects	Variance	SE	Standardized coefficient
Intercept variance (u_{oj})	299.40**	94.68	0.48
Level 1 variance (r_{ij})	2242.61***	86.15	0.54

* $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$

Note: Model controls not tabled. Level 1 controls included child gender, age, third grade math SOL scores, free/reduced lunch status, and ELL status. Level 2 controls included treatment assignment, masters degree, years teaching experience, school percent free/reduced lunch, school size, and school percent ELL. CPTS = Classroom Practices Teacher Survey; CPFS = Classroom Practices Frequency Survey; CPOM = Classroom Practices Observation Measure; MM = Morning Meeting; AC = Academic Choice; IM = Interactive Modeling