Abstract Title Page

Title:

Coaching Quality and Teachers' Implementation of the 4Rs Social-Emotional and Literacy Curriculum: Testing the Link between Two Levels of Intervention Fidelity

Authors and Affiliations:

Jason Downer (contact)
Center for Advanced Study of Teaching and Learning
University of Virginia
jdowner@virginia.edu

Josh Brown
Fordham University
cjobrown@fordham.edu

Manuela Jimenez Herrera
Center for Advanced Study of Teaching and Learning
University of Virginia
mj2qd@virginia.edu

Megan Stuhlman
Center for Advanced Study of Teaching and Learning
University of Virginia
mwstuhlman@gmail.com

Kyle Bourassa
Center for Advanced Study of Teaching and Learning
University of Virginia
kjb5x@virginia.edu

Ben Gologor
Fordham University
bgologor@fordham.edu

Pamela Wong
Harvard University
paw036@mail.harvard.edu

Abstract Body

Background / Context:

Studies from K-12 curricular research (O'Donnell, 2008) as well as school-based prevention (Greenberg, Domitrovich, Graczyk, & Zins, 2005) suggest that curricula implemented without a high degree of fidelity will fail to produce intended benefits. The importance of examining implementation quality during the evaluation of interventions has been recognized for decades (Carroll et al., 2007; Moncher & Prinz, 1991; Sechrest, West, Phillips, Redner, & Yeaton, 1979), yet there tend to be strikingly high rates of variability in implementation for past and recent trials of school-based program effectiveness (Berman & McLaughlin, 1976; Rohrbach, Grana, Sussman, & Valente, 2006). This is particularly important given Domitrovich and Greenberg's (2000) review of several social-emotional learning (SEL) programs, which found that implementation fidelity was closely linked to positive program outcomes for teachers and students. Thus, Chen (1998) and others (Devaney, O'Brien, Resnik, Keister, & Weissberg, 2006) suggest that the system in place for implementing interventions may be as important as the intervention itself in contributing to program outcomes.

Teacher-educators and policy-makers recognize that ongoing training and support for high quality implementation of curricula can be a vital component of systems that ensure the value of education experiences, particularly for students at-risk of school failure (Meisels, 2007; Pew Charitable Trusts, 2007; Pianta, 2005). In particular, there is growing consensus that one-on-one consultation approaches (also referred to as coaching, mentoring, staff development) that provide ongoing support and feedback may be the most direct, effective path to producing high quality implementation of curricula (Ingersoll & Kralik, 2004; Landry et al., 2006; Pianta & Allen, 2009). However, there is as much potential for variability in the quality of coaching as there is in a teacher's fidelity to a curriculum, which raises several questions about how these different levels of intervention fidelity are related to one another.

Purpose / Objective / Research Question / Focus of Study:

The current study is embedded within the evaluation of an integrated SEL and literacy program in third through fifth grade classrooms called the 4Rs (Reading, Writing, Respect, and Resolution; Morningside Center for Teaching Social Responsibility, 2001). A recent randomized controlled evaluation of 4Rs indicated moderate to large, positive impacts on teacher practices and small, positive impacts on student outcomes (Brown, Jones, LaRusso & Aber, 2010; Jones, Brown & Aber, 2011). Despite the promise of these findings, there was considerable variability during implementation of the 4Rs Program in curriculum delivery, which may have attenuated intervention effects. Therefore, in the current study teachers using the 4Rs curriculum are supported with a standardized, video-based coaching program called MyTeachingPartner (MTP; Pianta et al., 2008). In past trials with a variety of curricula across prek and secondary classrooms, MTP yielded improvements on teacher implementation and teaching practices that were moderate to large (Pianta et al., 2008).

As part of this evaluation of 4Rs+MTP, we are closely and objectively monitoring the quality with which coaches implement the 2 main components of MTP: (1) on-line written prompts that ask teachers to analyze key interactions from their own videotaped 4Rs lessons, and (2) in-person conferences that deepen teachers analysis of their 4Rs practice and result in

actionable improvement plans. In addition, videotaped footage of teachers' implementation of 4Rs are coded by independent observers for fidelity to the curriculum protocol. These data allow us to explore the following questions about patterns of variability in these 2 levels of implementation and how they relate to one another:

1. To what extent are the two measures of coaches' quality of implementation (prompt-writing, conferences) related to one another?
2. To what extent is there variability on these two measures within coaches, across their caseload of teachers?
3. To what extent does quality of coaching predict variation in teachers' implementation of the 4Rs curriculum, above and beyond a global measure of teaching quality (CLASS)?

Setting:

The 4Rs+MTP intervention is being implemented in 6 New York City public elementary schools.

Population / Participants / Subjects:

Across the 6 schools, 35 teachers are participating from grades 3-5. This sample includes 18 regular education classrooms, 3 special education classrooms, 10 inclusion classrooms (mix of regular and special education students) and 4 dual language classrooms. Schools and teachers are distributed between 2 4Rs coaches, with one coach serving 3 schools and 16 teachers and the other coach serving 4 schools and 19 teachers.

Intervention / Program / Practice:

The 4Rs program has two primary components: (1) a comprehensive 7-unit, 21 lesson, literacy-based curriculum in conflict resolution and social-emotional learning and (2) 25 hours of teacher training, plus ongoing coaching. The 4Rs curriculum uses high quality children's literature as a springboard for helping students gain skills and understanding in handling anger, listening, assertiveness, cooperation, negotiation, mediation, and building community. Each unit begins with the teacher reading aloud a children's book, carefully chosen to introduce the main theme of the unit. After the read-aloud is "Book Talk," which includes discussion, writing, and role-playing aimed at deepening students' understanding of the book. The rest of the unit consists of "Applied Learning" lessons in conflict resolution and social-emotional skills related to the theme. Lessons are taught every week for 30-45 minutes throughout the academic year.

MTP is an innovative approach to supporting curriculum implementation that unites the ideas of providing ongoing, personalized feedback and support to teachers and embedding these implementation supports within a validated framework that emphasizes the importance of teacher-student interactions to ensure effectiveness of curricula. The premise of MTP is that professional development for teachers can improve the implementation of curricula through provision of extensive *opportunities for individualized feedback* and support for effectiveness in one's own instruction, implementation, and interactions with students. Importantly, the Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Hamre, 2008), a shared framework for defining and observing teacher-student interactions that has shown strong links to growth in student outcomes, guides all observations, feedback, and support. Every two weeks,

teachers videotape their implementation of instructional activities and send this footage to their coach. The coach then edits the tape into short segments that focus on a specific dimension of interaction, as defined by the CLASS (e.g., teacher sensitivity), and its application to implementation of the instructional activity. The short video segments and accompanying written feedback and questions (called prompts) are posted to each teacher's private website. Teachers view the segments and respond to the accompanying prompts. Based on a long history of similar work focused on parent-child interactions (Dishion, Shaw, Connell, Gardner, Weaver, & Wilson, 2008; Juffer, Bakermans-Kranenburg, & van IJzendoorn, 2008; Mendelsohn et al., 2007; Webster-Stratton & Taylor, 2001), the intention of these prompts is to focus the teacher's attention on specific aspects of her behavior toward students and the students' responses. Teachers then meet with their coach for an in-person 30-minute conference to discuss the prompts and feedback, the teachers' responses, and to problem-solve and plan for future lessons. In this study, teachers will engage in 8 MTP coaching cycles throughout the academic year.

Research Design:

The study is a quasi-experimental, matched control group design. However, the current questions involve within-treatment group analyses.

Data Collection and Analysis:

Data on the *quality of coaches' online-prompts* to teachers will be assessed based on the written record of coach prompts stored and accessible to our research team via each teacher's individual website (3 prompts across 4 different coaching cycles = 12 prompts/teacher). The quality of each coaches online prompts to their teachers will be scored based on a coding rubric modified for use in this study. Eleven items, scored on a 3-pt. Likert scale (low quality=1, high quality=3), assess the coaches' prompts with regard to their specificity and detail in focusing teachers' attention on their interactions with students (e.g., CLASS dimension of focus is clearly defined; prompts are succinct, focused, and identify details about interactions with specific students) and the extent to which the coaches' questions/statements push teachers to reflect on and analyze those interactions and to consider something new about their practice.

Data on the *quality of in-person conferences* between coaches and teachers is being assessed using the transcripts of audiotaped coach-teacher conferences. The quality of 3-4 coaching conferences for each teacher will be scored based on a rubric developed for use in this study. Seventeen items, scored on a 3-point Likert scale (low quality=1, high quality=3), assess the extent to which in-person conferences deepen teachers' analysis of their 4Rs practice and result in actionable improvement plans. More specifically, items assess the degree to which coaches: achieve a balance between guiding the conference and promoting teacher expression, focus on specific teacher-student interactions, focus on the effects of teacher actions on students' learning and behavior, facilitate the teacher's ability to generate their own strategies for managing challenging situations, collaborate with teachers in identifying actionable next steps, and maintain a non-judgmental attitude towards the teacher's actions.

The *quality of teachers' implementation of 4Rs lessons* will be assessed using videotaped footage teachers take of themselves implementing 4Rs lessons in each of the first four coaching cycles of the year. Each full lesson videotape will be scored by independent observers based on a rubric developed specifically for this study to assess fidelity to the curriculum protocol.

Specifically, using a 5-pt. Likert scale (1=low, 5=high) that includes clear behavioral anchors characterizing each response category, observers rate the quality with which teachers conduct the read aloud (e.g., teacher elicits questions or comments about the book before reading; teacher reads book with expression/displays enjoyment of book), "Book Talk" (e.g., teacher ensures that dilemmas/choices in the story are clearly identified and discussed; teacher effectively uses discussion to explore the ideas expressed in student role plays), and "Applied Learning" (e.g., teacher effectively engages students in skill practice; the connection between the book reading and applied learning activities was made to the students).

Finally, the *global quality of teaching* will be measured by the CLASS during a single live observation in each classroom by an independent observer during the first month of the academic year. Observers rate teaching quality on a 7-pt. Likert scale (1=low, 7=high).

Bivariate correlations will be used to examine the extent to which quality of coaches' online-prompts are related to the quality of in-person conferences between coaches and teachers. Independent samples t-tests will be used to assess the extent of and differences in variability for each of the two quality measures within coaches, across their caseload of teachers. To examine the extent to which quality of coaching predicts variation in teachers' implementation of the 4Rs curriculum, above and beyond a global measure of teaching quality (CLASS), regression analyses will be conducted in Mplus (version 5.02, Muthen & Muthen, 2008) using the type=complex analysis command to account for the nesting of teachers within coaches.

Findings / Results:

Data for the proposed analyses using these measures in the current sample are being collected and coded this fall and will be available for analysis by January 2013. Important to note is that pilot testing last year of the 2 coaching quality assessment rubrics across 4 coaches and the 4Rs implementation quality measure across 12 demographically similar classrooms demonstrated a high level of agreement between coders. In addition, 20 videos were coded across these 12 classrooms for quality of 4Rs implementation, which indicated normal distribution of items and variability across classrooms (M=3.29, SD=.56); this provides confidence that there will be variability to predict in this measure across the new sample of 35 classrooms.

Conclusions:

We expect to learn how different measures of coaching implementation are associated with one another and to identify elements of coaching that may be particularly challenging to standardize. Implications will be drawn about how best to provide support to coaches in the field to maintain their fidelity to the coaching model and ensure a high quality coaching experience for teachers. In addition, we will discuss patterns of prediction from quality of coaching implementation to teachers' fidelity to the 4Rs intervention model in light of implications for identifying elements of coaching that might be most strongly linked to teachers' classroom practice. In particular, we will draw attention to the distinction between the on-line element of coaching (prompts) and the in-person aspect (conferences), given the significant logistical and resource implications of providing support to teachers remotely over the internet versus live in the classroom.

Appendices

Appendix A. References

Berman, P. & McLaughlin, M.W. (1976). Implementation of educational innovation. *Educational Forum, 40,* 347-370.

Brown, J.L., Jones, S.M., LaRusso, M.D., & Aber, J.L. (2010). Improving classroom quality: Teacher influences and experimental impacts of the 4Rs Program. *Journal of Educational Psychology*, *102*(1), 153-167.

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, *40*, 1-9.

Chen, H. (1998). Theory-driven evaluations. *Advances in Educational Productivity, 7,* 15–34.

Devaney, E., O'Brien, M. U., Resnik, H., Keister, S., & Weissberg, R. (2006). *Sustainable schoolwide social and emotional learning (SEL): Implementation guide and toolkit*. Chicago, IL: Collaborative for Academic, Social, and Emotional Learning (CASEL).

Dishion, T. J., Shaw, D. S., Connell, A. M., Garner, F., Weaver, C. M., & Wilson, M. N. (2008). The Family Check-Up with high-risk indigent families: Preventing program behavior by increasing parents' positive behavior support in early childhood. *Child Development, 79*(5), 1395-1414.

Domitrovich, C., & Greenberg, M. T. (2000). The study of implementation: Current findings from effective programs that prevent mental disorders in school-aged children. *Journal of Educational & Psychological Consultation*, *11*, 193–221.

Greenberg, M. T., Domitrovich, C. E., Graczyk, P. A., & Zins, J. E. (2005). The study of implementation in school-based preventive interventions: Theory, research, and practice. Promotion of mental health and prevention of mental and behavioral disorders, Volume 3. DHHS Pub. No. Rockville MD: Substance Abuse and Mental Health Services Administration.

Ingersoll, R., & Kralik, J. M. (2004). *The impact of mentoring on teacher retention: What the research says*. Denver: Education Commission of the States.

Jones, S.M., Brown J.L., & Aber, J.L. (2011). Two year impacts of a universal school-based social-emotional and literacy intervention: An experiment in translational developmental research. *Child Development, 82,* 533-554.

Juffer, F., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2008). *Promoting positive parenting: An attachment-based intervention*. New York: Lawrence Erlbaum Associates.

Landry, S.H., Swank, P.R., Smith, K.E., Assel, M.A., & Gunnewig, S.B. (2006). Enhancing early literacy skills for preschool children: Bringing a professional development model to scale. *Journal of Learning Disabilities*, 39(4), 306-324.

Meisels, S. J. (2007). Accountability in early childhood: No easy answers. In R. C. Pianta, M. J. Cox, & K. Snow (Eds.), *School readiness & the transition to kindergarten in the era of accountability* (pp. 49–84). Baltimore, MD: Brookes.

Mendelsohn, A. L., Valdez, P. T., Flynn, V., Foley, G. M., Berkule, S. B., Tomopoulos, S., Fierman, et al. (2007). Use of videotaped interactions during pediatric well-child care: Impact at 33 months on parenting and on child development. *Journal of Developmental & Behavioral Pediatrics*, *28*, 206-212.

Moncher, F. J., & Prinz, R. J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review, 11*, 247–266.

Morningside Center for Teaching Social Responsibility. (December, 2008). *Executive Summary of Research Report to WT Grant Foundation*. Unpublished report, Morningside Center for Teaching Social Responsibility.

Muthen, L. K., & Muthen, B. (2008). *Mplus user's guide.* Los Angeles CA: Muthen & Muthen.

Pew Charitable Trusts. (2005). National Early Childhood Accountability Task Force Briefing. Philadelphia: Pew.

Pianta, R. C. (2005). Standardized observation and professional development: A focus on individualized implementation and practices. In M. Zaslow and I. Martinez-Beck (Eds.), *Critical issues in early childhood professional development* (pp. 231-254). Baltimore: Paul H. Brookes Publishing.

Pianta, R., & Allen, J. (2009). Building capacity for positive youth development in secondary school classrooms: Changing teachers' interactions with students. In M. Shinn & H. Yoshikawa (Eds.), *Toward positive youth development: Transforming schools and community programs*. New York: Oxford University Press.

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System (CLASS) manual, K–3.* Baltimore, MD: Brookes.

Pianta, R. C., Mashburn, A. J., Downer, J., Hamre, B. K., & Justice, L. (2008). Effects of Web-mediated professional development resources on teacher-child interactions in pre-kindergarten classrooms. *Early Childhood Research Quarterly, 23,* 431–451.

O' Donnell, 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*, 33-84.

Rohrbach, L. A., Grana, R., Sussman, S., & Valente, T.W. (2006). Type II translation: Transporting prevention interventions from research to real-world settings. *Evaluation & the Health Professions*, *29*(3), 1–32.

Sechrest, L., West, S. G., Phillips, M. A., Redner, R., & Yeaton, W. (1979). Some neglected problems in evaluation research: Strength and integrity of treatments. In L. Sechrest and associates (Eds.), *Evaluation Studies Review Annual (Vol. 4)* (pp. 15-35). Beverly Hills, CA: Sage Publications.

Webster-Stratton, C., & Taylor, T. (2001). Nipping early risk factors in the bud: Preventing substance abuse, delinquency and violence in adolescence through interventions targeted at young children (0-8 years). *Prevention Science*, *2*(3), 165-192.

Appendix B. Tables and Figures

Not applicable.