Contrasting traditional, one-on-one teacher coaching versus coaching teacher pairs: An RCT testing impacts and sustainability

Elise T. Pas[1], Katrina J. Debnam[2], Chelsea A. Duran[2], Catherine P. Bradshaw[2]

[1] Johns Hopkins University
[2] University of Virginia

Introduction and aims. Tailored, data-driven teacher coaching shows promise for enhancing teachers' knowledge and skills, improving instructional practices, reducing use of discipline referrals for Black students, and reducing student disruption (Denton & Hasbrouck, 2009; Neuman & Cunningham, 2009; Pianta, Mashburn, et al., 2008; Reinke et al., 2008). Notably, much of the extant coaching research has focused on traditional, one-to-one coaching; the impact of coaching teacher pairs has been considered less often, despite the potential for greater feasibility/efficiency, as well as longer-sustained impacts through near-peer-support (Cappella et al., 2012; Wandersman et al., 2008).

Thus, this study tested a "teacher-to-teacher" (T2T) coaching model, whereby a coach worked with two teachers simultaneously, in a randomized controlled trial. Effects were compared to the effects ascertained in a prior trial testing a one-on-one coaching approach (Bradshaw et al., 2018). We sought to estimate: ***(Aim 1)*** the main effects of coaching (either as one-on-one and/or T2T) on teacher classroom management practices and student behavior immediately at post-test as well as one year following post-test; ***(Aim 2)*** differential effects between the coaching models; and ***(Aim 3)*** relative time efficiency between the coaching models.

Method. Participants were 152 teachers in 18 schools (9 elementary and 9 middle) recruited in two separate stages to participate in a two-armed study examining the effects of "Double Check", a professional development and coaching intervention promoting teachers' use of culturally responsive practices to improve student engagement. The first 12 schools were recruited into the one-on-one coaching arm of the study; an additional 6 schools were recruited into the T2T arm of the study. All teachers in all schools were expected to attend five "Double Check" professional development sessions. Coaching was provided only to the teachers who consented and were randomized (at the teacher-level) to intervention status. The coaching model was an adaptation of the Classroom Check-Up (Reinke, 2006; Reinke et al., 2011; Reinke et al., 2008), involving interviews, feedback, and goal setting with the coach. In the T2T study arm, these activities were conducted with both teachers simultaneously, and each teacher also observed their paired peer's classroom.

Data were collected at the beginning of the first school year of participation (i.e., pre-test; T1), at the end of the first school year of participation (i.e., at post-test; T2), and again one year following post-test (i.e., T3). Teachers completed online surveys and reported about their perceived self-efficacy and work-related stress on 5-point Likert scale items. Perceived self-efficacy was assessed with three separate scales: (1) the *Multicultural Efficacy Scale* (Guyton & Wesche, 2005); (2) the *Culturally-Responsive Teaching Self-Efficacy Scale* (Siwatu, 2007); and (3) Classroom behavior management self-efficacy using the Hoy and Woolfolk (1993) efficacy scale. External observers conducted a single, roughly 15-minute classroom observation of each teacher at each time point using the Assessing School Settings: Interactions of Students and Teachers (ASSIST; Rusby, Crowley, Sprague, & Biglan, 2011; Rusby, Taylor, & Milchak, 2001). ASSIST measures included event-based tallies as well as global ratings (comprising 5-point Likert scale items) of student and teacher behaviors. The School-Wide Information System (Irvin et al., 2004; May et al., 2003) was used to track office discipline referrals (ODRs).

We conducted three-level, hierarchical linear models using HLM 7 software (Raudenbush, Bryk, & Congden, 2013), with observations (Level 1) nested within teachers (Level 2), who were nested within schools (Level 3). Nonlinear change over the three time points was modeled using dummy codes for T2 and T3. Models included study arm (i.e., T2T arm = 1; one-on-one arm = 0) as a covariate. Cross-level interactions between study arm and the T2 and T3 dummy variables were included to account for differences in trends observed between control

groups of the study arms. Other controls in the analyses were: teachers' social desirability bias (teacher self-report), teacher race/ethnicity, age, and gender (self-report), as well as school level (elementary vs. middle school).

Results (see Table 1). ***Aim 1.*** Model 1 included intervention condition (coaching versus not coached) as a predictor of the Level 1 coefficients for the T2 and T3 dummies, thereby modeling treatment effects independently at each time point. The results across the 20 outcomes were mixed. At T2, three outcomes suggested desirable effects (i.e., tallies of teacher use of approvals, tallies of student non-cooperation, and global ratings of teachers' behavior management) and one suggested an undesirable effect (tallies of teacher use of disapprovals). There was also an undesirable treatment effect at T3 on ODRs for White students, whereby the rates of ODRs for White students was elevated for coached teachers relative to not-coached teachers.

Aim 2. In order to address differential effects by study arm, Model 2 included a cross-level interaction between study arm and the intervention effects at each time point. The impacts of the one-on-one coaching modeled here at T2 reproduce those reported in Bradshaw et al. (2018). The T2T coaching was less effective than one-on-one coaching in increasing student cooperation and decreasing student disruptive behaviors. Analyses positing the T2T arm sub-sample as reference group yielded no significant effects for T2T coaching at T2. At T3, a differential treatment effect was observed in ASSIST tallies of student non-cooperation, reflecting an undesirable, significant treatment effect of T2T coaching on student non-cooperation.

Aim 3. Coaches in the T2T coaching study spent an average of 446.9 minutes per teacher (*SD* = 139.2 minutes), whereas coaches in the one-on-one coaching study arm spent an average of 522.3 minutes per teacher (*SD* = 195.1 minutes; difference: 75.4 minutes, $p < .01$). The two coaching models did not differ in terms of teachers' time spent with coaches ($p = .49$).

Discussion. While demonstrating the feasibility of potential time-savings for coaches, this particular, small-scale trial suggests there is no such time-saving for teachers and the T2T model does not replicate the effects of a coaching intervention as when delivered in a one-on-one format. Further investigation and analyses are needed to explore additional implementation and experiential factors differing between coaching in one-on-one and paired formats.

References

Bradshaw, C. P., Pas, E. T., Bottiani, J. H., Debnam, K. J., Reinke, W. M., Herman, K. C., & Rosenberg, M. S. (2018). Promoting cultural responsivity and student engagement through Double Check coaching of classroom teachers: An efficacy study. *School Psychology Review*, *47*(2), 118-134.

Cappella, E., Hamre, B. K., Kim, H. Y., Henry, D. B., Frazier, S. L., Atkins, M. S., & Schoenwald, S. K. (2012). Teacher consultation and coaching within mental health practice: Classroom and child effects in urban elementary schools. *Journal of consulting and clinical psychology*, *80*(4), 597.

Denton, C. A., & Hasbrouck, J. A. N. (2009). A description of instructional coaching and its relationship to consultation. *Journal of Educational and Psychological Consultation*, *19*(2), 150-175.

Guyton, E. M., & Wesche, M. V. (2005). The multicultural efficacy scale: Development, item selection, and reliability. *Multicultural Perspectives*, *7*(4), 21–29.

Hoy, W. K., & Woolfolk A. E., (1993). Teachers' sense of efficacy and the organizational health of schools. *The Elementary School Journal*, *93*(4), 355–372.

Irvin, L., Tobin, T., Sprague, J., Sugai, G., & Vincent, C. (2004). Validity of office discipline referral measures as indices of school-wide behavioral status and effects of school-wide behavioral interventions. *Journal of Positive Behavior Interventions*, *6*, 131–147.

May, S., Ard, W., Todd, A. W., Horner, R. H., Glasgow, A., Sugai, G., & Sprague, J. (2003). *Schoolwide Information System*. Eugene, OR: University of Oregon, Educational & Community Supports.

Neuman, S. B., & Cunningham, L. (2009). The impact of professional development and coaching on early language and literacy instructional practices. *American educational research journal*, *46*(2), 532-566.

Pianta, R. C., Mashburn, A. J., Downer, J. T., 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*(4), 431-451.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R., & du Toit, M. (2011). *HLM 7: Hierarchical linear and nonlinear modeling*. Skokie, IL: Scientific Software International.

Reinke, W. M. (2006). The Classroom Check-Up: Advanced tools for effective coaching. In R. Sprick, J. Knight, W. M. Reinke, & T. McKale, (Eds.), *Coaching classroom management: Strategies and tools for administrators and coaches* (pp. 139–167). Eugene, OR: Pacific Northwest Publishing.

Reinke, W. M., Herman, K. C., & Sprick, R. (2011). *Motivational interviewing for effective classroom management: The Classroom Check-Up*. New York, NY: Guilford Press.

Reinke, W. M., Lewis-Palmer, T., & Merrell, K. (2008). The Classroom Check-Up: A classwide teacher consultation model for increasing praise and decreasing disruptive behavior. *School Psychology Review, 37,* 315–332.

Rusby, J. C., Crowley, R., Sprague, J., & Biglan, A. (2011). Observations of the middle school environment: The context for student behavior beyond the classroom. *Psychology in the Schools, 48,* 400–415.

Rusby, J. C., Taylor, T., & Milchak, C. (2001). *Assessing school settings: Interactions of students and teachers (ASSIST) observation system*. Unpublished manuscript. Eugene, OR: Oregon Research Institute.

Siwatu, K. O. (2007). Preservice teachers' culturally responsive teaching self-efficacy and outcome expectancy beliefs. *Teaching and Teacher Education, 23*(7), 1086–1101.

Wandersman, A., Duffy, J., Flaspohler, P., Noonan, R., Lubell, K., Stillman, L., ... & Saul, J. (2008). Bridging the gap between prevention research and practice: The interactive systems framework for dissemination and implementation. *American journal of community psychology*, *41*(3-4), 171-181.

Table 1 – Effect Estimates for Combined/Main Effects and Differential Effects at Post-Test and One-Year Follow-Up.

	Model 1				Model 2							
	Combined Effects				Treatment Effect (Traditional Coaching Model)				Differential Treatment Effects (T2T vs. Traditional Coaching)			
	T2 (post-test)		T3 (follow-up)		T2		T3		T2		T3	
Disciplinary Referrals	Incident Rate Ratios (95% Confidence Intervals)											
Total	0.89	(0.70,1.12)	0.96	(0.74,1.25)	0.85	(0.65, 1.10)	0.84	(0.57, 1.24)	1.22	(0.68, 2.19)	1.29	(0.76, 2.18)
Black Students	0.78	(0.57,1.05)	0.89	(0.59,1.36)	0.71*	(0.52, 0.98)	0.82	(0.47, 1.43)	2.13	(0.79, 5.74)	1.26	(0.54, 2.91)
White Students	1.09	(0.69,1.70)	1.61*	(1.02,2.56)	1.35	(0.80, 2.29)	0.88	(0.37, 2.10)	0.38	(0.13, 1.12)	2.11	(0.76, 5.86)
ASSIST Tallies	Incident Rate Ratios (95% Confidence Intervals)											
Proactives	1.05	(0.89,1.24)	1.04	(0.88,1.22)	1.04	(0.83, 1.31)	0.98	(0.78, 1.23)	1.02	(0.73, 1.43)	1.12	(0.81, 1.55)
Approvals	1.33*	(1.04,1.70)	1.19	(0.92,1.55)	1.51*	(1.09, 2.10)	1.33	(0.96, 1.84)	0.74	(0.44, 1.22)	0.73	(0.42, 1.28)
Disapprovals	1.82*	(1.00,3.32)	1.60	(0.91,2.83)	2.08	(0.99, 4.37)	1.04	(0.49, 2.21)	0.68	(0.19, 2.40)	2.64	(0.80, 8.66)
OTR	1.03	(0.87,1.21)	1.02	(0.84,1.23)	0.98	(0.79, 1.22)	1.02	(0.81, 1.28)	1.11	(0.80, 1.54)	0.99	(0.66, 1.47)
Student Noncooperation	0.53*	(0.31,0.88)	1.22	(0.79,1.9)	0.52	(0.26, 1.03)	0.66	(0.37, 1.19)	0.95	(0.34, 2.68)	4.11*	(1.65, 0.25)
Student Disruptives	0.97	(0.78,1.20)	1.25	(0.98,1.59)	0.97	(0.74, 1.27)	1.03	(0.75, 1.41)	1.00	(0.64, 1.57)	1.54	(0.95, 2.49)
ASSIST Global Ratings	Spybrook's Delta (SE)											
CR Practice	0.04	(0.12)	0.01	(0.13)	0.18	(0.17)	-0.1	(0.19)	-.36	(0.27)	.22	(0.3)
Control	0.07	(0.07)	-0.11	(0.08)	0.26	(0.17)	-0.1	(0.19)	-.33	(0.28)	-.28	(0.3)
Anticipation	0.18+	(0.10)	-0.08	(0.11)	0.41	(0.16)	-0.1	(0.19)	-.41	(0.27)	-.01	(0.28)
Beh. Mgmt.	0.29***	(0.08)	-0.10	(0.09)	0.56***	(0.17)	-0.3	(0.18)	-.29	(0.26)	.26	(0.29)
Meaningful Part.	0.06	(0.1)	-0.07	(0.11)	0.12	(0.15)	-0.1	(0.17)	-.12	(0.24)	.00	(0.27)
Student Compliance	0.13	(0.09)	-0.13	(0.10)	0.43*	(0.16)	-0.1	(0.19)	-.62*	(0.26)	-.29	(0.29)
Student Disruptive Behaviors	-0.11+	(0.06)	0.06	(0.07)	-0.4**	(0.16)	-0.1	(0.18)	.53*	(0.27)	.51+	(0.29)
Teacher Self-Reports	Spybrook's Delta (SE)											
CRT Efficacy	-0.04	(0.06)	-0.05	(0.07)	0.00	(0.13)	-0.1	(0.19)	-.15	(0.23)	-.06	(0.28)
Multicult. Efficacy	0.00	(0.05)	-0.04	(0.07)	0.00	(0.13)	-0.1	(0.20)	.00	(0.22)	.02	(0.28)
Beh. Mgmt. Efficacy	0.01	(0.07)	-0.12	(0.10)	0.12	(0.14)	-0.1	(0.20)	-.29	(0.23)	-.28	(0.31)
Stress	-0.12	(0.11)	0.16	(0.15)	-0.17	(0.14)	0.11	(0.19)	.16	(0.22)	.11	(0.29)

NOTES: T2 = Time 2 (end-of-year post-test); T3 = Time 3 (one-year follow-up); OTR = Opportunities to Respond; CR = Culturally Responsive; Beh. Mgmt. = Behavior Management; Meaningful Part. = Meaningful Participation; CRT = Culturally Responsive Teaching; Multicult. = Multicultural.
***$p < .001$, **$p < .01$, *$p < .05$, +$p < .10$.