


Type and Dosage of Performance Feedback Following COMPASS Consultation on Teacher and Student Outcomes

Remedial and Special Education
2024, Vol. 45(1) 30–43
© Hammill Institute on Disabilities 2023
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/07419325231164755
rase.sagepub.com



Lindsey N. Ogle, PhD¹ , Lisa A. Ruble, PhD¹ ,
Michael D. Toland, PhD², and John H. McGrew, PhD³ 

Abstract

Little is known about differences in dosage or feedback that make teacher coaching effective in relation to teacher and student outcomes. This study builds upon previous research on the Collaborative Model for Promoting Competence and Success (COMPASS) to understand the impact of different types (face-to-face coaching vs. emailed feedback) and dosages (one vs. two or four) of performance feedback compared to receiving no feedback following an initial consultation during which intervention plans were developed. Findings suggest that teacher adherence and student goal-attainment outcomes depend on dosage, not type of coaching. Specifically, having two or four opportunities for performance feedback was significantly better than having none or only one opportunity. Ratings of teacher adherence and student goal attainment delivered via emailed feedback or face-to-face coaching were similar, which may have important implications for cost and efficiency. While preliminary, the results are promising and warrant further research. Implications are also discussed.

Keywords

collaboration/consultation, evidence-based practice, autism, personnel preparation: effective practice, family/parental involvement

As opposed to traditional methods of professional development (PD) that often fail to include features that support implementation of classroom interventions, coaching is a research-supported method for improving teacher instruction and student outcomes (Dunst et al., 2015; Kraft et al., 2018; Kretlow & Bartholomew, 2010; Morrier et al., 2011; Ruble et al., 2010, 2012, 2013a, 2018, Ruble & McGrew, 2015). Coaching, while individualized to the specific needs of the teacher, often focuses on discrete skills and incorporates performance feedback (PF) on teaching quality and student progress over a longer period (Beidas et al., 2012; Kraft et al., 2018; Kretlow & Bartholomew, 2010). Combined with high-quality intervention plan development, coaching with PF is an effective method for helping teachers implement intervention programs for students with complex learning needs such as learners with autism spectrum disorder (ASD). Research shows that when teachers of students with ASD receive coaching, improved student outcomes and fidelity to the teaching plan/intervention are observed (Brock et al., 2020; Hamrick et al., 2021; Ruble et al., 2010, 2012, 2013a, 2015, 2018).

But what makes coaching so effective? Is it the amount of coaching, a specific component of coaching (i.e., PF), or

are both important? Recent research conducted on coaching has found conflicting results. In a meta-analysis of 60 studies of coaching efficacy, Kraft et al. (2018) reported moderate improvements in teacher's quality of instruction ($d = 0.49$) and modest improvements in student achievement ($d = 0.18$) with coaching. However, coaching dosage varied widely, with 27% of studies reporting 10 hours or less of one-to-one coaching, 23% reporting 11 to 20 hours, and 23% reporting 21 or more hours. Despite these wide variations in the amount of coaching received, they found no consistent association among coaching dosage, quality of teacher instruction, and student achievement. In another recent meta-analysis, Brock and Carter (2017) also reported a lack of a relationship between duration of training and quality of

¹Ball State University, Muncie, IN, USA

²University of Toledo, OH, USA

³Indiana University-Purdue University, Indianapolis, USA

Corresponding Author:

Lindsey N. Ogle, Department of Special Education, Teachers College
Room 705, Ball State University, Muncie, IN 47306, USA.
Email: lno@bsu.edu

Associate Editor: Wendy Oakes

teacher instruction (i.e., implementation fidelity). These findings suggest that dosage, defined in terms of duration or amount, may not be as important as what occurs during coaching (i.e., the content and/or quality of coaching).

However, a recent meta-analysis of 71 studies of PF in multiple settings (e.g., human service organizations, schools, retail stores, restaurants) conducted by Sleiman et al. (2020) found that dosage does matter. Across these varied settings, results showed that the frequency of PF (daily or weekly) was associated with large to very large effect sizes (ES) greater than $d = 0.6$ for 81% of the studies ($M_d = 0.78$; $n = 96$; large ES). More specifically, they found that PF mostly in positive verbal, written, and graphical formats with goal setting and behavioral consequences (e.g., praise) was associated with the largest ES (Sleiman et al., 2020). These findings are consistent with educational research in which coaching with PF is an established evidence based practice (EBP) often applied within a problem-solving consultation framework (Fallon et al., 2015). In a meta-analysis of 29 studies of treatment plan implementation in schools, Noell et al. (2014) found that follow-up sessions after an initial consultation without a review of treatment implementation or student outcomes were ineffective for improving intervention implementation, whereas PF and self-monitoring with environmental supports were highly effective in improving implementation and student outcomes.

Implementation science is a relatively new but important focus of research in ASD. For example, with respect to ASD implementation research, studies of PF in coaching are emerging and can be divided into two broad areas: (a) assessment of implementation fidelity of focused EBPs and (b) implementation of intervention plans based on common elements or principles that are features of multiple EBPs. For the former, examples are teacher-delivered behavior-specific praise (Ennis et al., 2020), discrete trial teaching (Fraser et al., 2020), and the implementation of focused EBPs outlined in the AFIRM modules (AFIRM, Sam et al., 2021; <https://afirm.fpg.unc.edu/node/137>) for students with ASD (Mahoney, 2020; Sam et al., 2021). In their intervention, Sam et al. (2021) coached teachers to implement a focused EBP chosen by the teacher for a specific student with ASD. The coaching process included completing an online training module on the focused EBP, receiving coaching that consisted of a meeting prior to an observation, observation of the implementation fidelity of the focused EBP, and postobservation debriefs reviewing results from a fidelity checklist. The results showed that coaching resulted in teachers not only using more EBPs but also implementing those EBPs with higher implementation fidelity, which in turn resulted in higher student goal-attainment outcomes than for students whose teachers had implemented services as usual (Sam et al., 2021). An example illustrating the latter broad area, Ruble et al. (2010, 2012, 2013a, 2015, 2018,

2022) conducted a series of randomized controlled trials (RCTs) examining a personalized parent-teacher goal setting, intervention planning, and coaching intervention based on shared decision-making called Collaborative Model for Promoting Competence and Success (COMPASS). Rather than focusing on the implementation of focused EBPs, COMPASS coaching focuses on teacher implementation of individualized teaching plans guided by selected EBPs personalized for the specific needs and resources of the student (Ruble et al., 2022) together with an approach that incorporates high-leverage practices (McLeskey et al., 2017) and applies principles common to high-quality intervention plans (Ruble et al., 2020, 2022). The distinction between the content of the two different approaches is that the former addresses PF as part of a focused EBP while the latter provides PF of multiple adapted EBPs to the specific student and their situation (Ruble et al., 2020, 2022).

The Collaborative Model for Promoting Competence and Success

As a consultation and coaching intervention, COMPASS (Ruble et al., 2012) embeds goal setting, intervention planning, and coaching with PF on adherence to intervention plans and student goal progress. The Collaborative Model for Promoting Competence and Success is manualized, has been successfully tested in three RCTs (Ruble et al., 2010, 2013a, 2018) for students with ASD across preschool and high school ages and has produced consistently high ES for improved student outcomes ($d = 1.1$ – 2.0).

Initial consultation. The Collaborative Model for Promoting Competence and Success begins with an initial consultation with the student's caregiver, teacher, and other relevant school specialists (e.g., speech-language pathologists), which results in three individualized social-emotional learning (SEL) goals and intervention plans using an evidence-based practice in psychology framework (Ruble & McGrew, 2015). With consultant support, the COMPASS profile that describes the student's strengths and challenges in various domains (e.g., communication skills, social skills, challenging behaviors, independent learning skills) from parent and teacher inputs are reviewed. From this in-depth discussion, three SEL goals in the domains of social, communication, and independent learning skills are identified based on National Research Council (2001) recommendations. The remaining time focuses on the collaborative development of intervention plans for each goal with a detailed step-by-step teaching plan utilizing a common elements approach for instruction (Ruble et al., 2021) and the appropriate EBPs related to the goal (AFIRM, Sam et al., 2020; <https://afirm.fpg.unc.edu/node/137>) adapted to the child's strengths and needs (Ruble et al., 2022).

Coaching. After this initial consultation in standard COMPASS, teacher coaching is provided to support the implementation of the intervention plans (Ruble et al., 2012) with four 1-hour sessions that occur every 4 to 6 weeks. Coaching sessions begin with a review of the teacher's implementation of each of the intervention plans as captured by a teacher-made video. Afterward, the teacher and consultant discuss the student's goal-attainment progress and the teacher's adherence to the intervention plans, making any adaptations or updates as needed. In addition to PF on adherence to the intervention plans and student goal-attainment progress, COMPASS coaching also incorporates video self-reflection—another established EBP in training teachers (Morin et al., 2019; Nagro & Cornelius, 2013). Reviewing videos allows teachers to self-reflect on their own practice and problem-solve issues they are having with implementing the intervention plans with their coach in a way that would not be possible during live instruction with a student with autism (Morin et al., 2019). The targeted PF on both intervention practice and student outcomes via video analysis in coaching has been shown in previous studies of COMPASS to be highly impactful on teacher's adherence to the intervention plans and student goal-attainment outcomes (Ruble et al., 2013a). For instance, Ruble et al. (2013a) demonstrated that teacher adherence increased significantly over four coaching sessions and, more importantly, that teacher adherence was associated with student goal-attainment outcomes (Wong et al., 2018). Similar results have been found for novice vs. veteran teachers and for students with high support needs vs. lower support needs (Ruble et al., 2010, 2013a, 2018), suggesting that COMPASS is an intervention that is highly adaptable to the specific needs of the teacher and student. In total, the consultant provides approximately 7 hours of direct consultation/coaching support to teachers and caregivers, which is a relatively low dosage compared to other coaching interventions (Kraft et al., 2018).

Current Study

However, questions remain about the specific mechanisms that underpin the effectiveness of COMPASS coaching. To date, outcomes of COMPASS have been evaluated based on four teacher coaching sessions delivered face-to-face or via video conferencing technology. However, the extent to which student outcomes might vary depending on the dosage of coaching is unknown.

Based on previous research on COMPASS, whether PF on the teacher's adherence to the intervention plans and the student's goal-attainment progress alone without the added benefits of problem-solving issues with implementation with a coach is enough to demonstrate high teacher adherence and student outcomes is not clear. To this aim, the study design included a condition in which teachers received emailed PF reports (e-feedback) on their adherence to the

intervention plans and student goal-attainment progress along with limited suggestions for improvement and encouraging comments. Just like for coaching, PF was based on teacher-made videos of the implementation of each of the intervention plans and progress-monitoring data. While e-feedback does not allow for interpersonal problem solving with the coach, it is much more efficient and cost-effective than coaching. E-feedback takes on average about a half hour to complete based on teacher-provided videos compared to roughly an hour and a half for face-to-face coaching to conduct the coaching session and complete and email the report. In addition, e-feedback can be reviewed when convenient and does not require scheduling, travel, or time away from the classroom for the teacher, which can be costly.

Last, a condition in which teachers and caregivers only received an initial consultation in which the student's intervention plan was developed without any follow-up coaching or e-feedback was included. Based on previous research on COMPASS, having a high-quality intervention plan individualized to the specific needs and learning contexts of the student is essential. Thus, all participants in the current study received an initial COMPASS consultation. In this condition, teachers were instructed to implement the intervention plans as written, but consultant trainees (CTs) were not permitted to discuss the intervention plans with these teachers or caregivers following the initial consultation. This condition served as control to compare the different dosages and types of PF.

Thus, our research questions concerned dosage and type (i.e., coaching vs. e-feedback) of PF on teacher adherence and student outcomes. We had the following research questions: (RQ1) Do teachers who receive more PF (dosage) have higher child goal-attainment outcomes and better adherence to the intervention plans? (RQ2) Does the type of PF (coaching vs. e-feedback) impact child goal-attainment outcomes and teacher adherence to the intervention plans? Due to differences in when the initial consultation occurred in the current study, an exploratory question concerned the onset of the intervention and whether consultations conducted earlier in the school year resulted in better outcomes.

Method

Participants

Participants were recruited in a multistep fashion from a mid-southern U.S. state across two school years. During the second year, schools began shutting down because of COVID-19. All participants were recruited from public schools. Special education directors were contacted first by the researchers via email for permission to conduct research in their schools and train one of their ASD consultants in COMPASS. Once they agreed to participate,

directors shared the contact information for the consultant, who were recruited over the phone by the researchers. In total, 60 special education directors were contacted, of which 12 responded with interest in participating. Twelve consultants, one from each school corporation, indicated initial interest in participating, but three declined to participate due to personal issues prior to being trained. No consultants dropped out from the study once trained.

The Collaborative Model for Promoting Competence and Success consultants were trainees (CTs) naïve to the COMPASS intervention, which is different from all prior RCTs of COMPASS (Ruble et al., 2010, 2013, 2018) where developers served as the sole consultants. To test whether COMPASS-naïve individuals could acquire the critical COMPASS skills in consultation and coaching, a training package was developed based on stakeholder input to train school-based consultants to implement COMPASS with high multidimensional fidelity (Ruble et al., 2022). Ruble et al. (2022) evaluated the training package and confirmed highly positive implementation outcomes when COMPASS was delivered by school- or community-based trainees.

CTs were eligible to participate if they provided school consultation or training to teachers of students with ASD, were not in a supervisory role over teachers in the study, and did not intend to leave their jobs during the school year. In total, nine CTs were recruited. All CTs were White females between the ages of 24 and 47 years ($M = 37$, $SD = 7.96$) with average consultation experience of 10 years ($SD = 6.31$). Eight CTs had a master's degree, one had a bachelor's degree, and seven had prior teaching experience. Seven trainees were from school systems, and two were advanced doctoral students in psychology (different from the researchers' university) who consulted with schools via rural ASD clinics.

After the CTs were recruited, they worked with special education directors to identify teachers of students with ASD. Special education teachers were eligible to participate if they had at least one student with ASD on their caseload, agreed to the study activities, and had no intention to leave during the school year. After teachers were recruited, a student with ASD was randomly chosen from their caseload to participate, and researchers contacted the students' caregivers after caregivers provided permission to be contacted by the researchers. To be eligible for study participation, all students needed to be receiving special education services under the category of ASD and meet cutoff scores for ASD with the Social Communication Questionnaire–Current (SCQ-C; Rutter et al., 2003). In total, 28 sets of teachers, caregivers, and students participated, and their demographics are summarized in Table 1. CTs each worked with two to four teachers, and two of the 28 sets (7%) of participants were in the same school.

Procedure

CT training and supervision. CTs were provided with training on COMPASS by the researchers. The training consisted of online training modules on CANVAS Free for Teachers (<https://www.instructure.com/canvas/login/free-for-teacher>) and two in-person 6-hour training days delivered by the researchers that focused on the initial consultation and on coaching/emailed PF, respectively. The two in-person training sessions were spaced roughly a month apart to give CTs the opportunity to conduct at least one initial consultation prior to the coaching/emailed feedback training. CTs received supervision from the researchers following each consultation, coaching session, and emailed feedback report focusing on the CTs' fidelity to implementation, quality of delivery, and teacher/caregiver acceptance (Ruble et al., 2022). In total, CTs received approximately 28 hours of direct training by the researchers in COMPASS consisting of 12 hours of in-person training, approximately 12 hours of online self-directed training, and four 1-hour supervision sessions (two consultation and two coaching), followed by additional supervision reports for each consultation and coaching session conducted by the CT. Ruble et al. (2022) present the implementation outcomes of the COMPASS training package in further detail. Overall, the training package was found to be effective in training CTs to implement COMPASS with high fidelity and high teacher/caregiver responsiveness (Ruble et al., 2021, 2022).

Initial consultation. All teachers and caregivers received an initial COMPASS consultation facilitated by a COMPASS-trained CT between October and January. During the 3-hour initial consultation, the student's COMPASS profile was discussed, and three goals in the domains of social skills, communication skills, and independent learning skills were selected based on National Research Council (2001) recommendations. Intervention plans to meet each goal were then collaboratively developed based on the student's personal and environmental challenges and supports incorporating step-by-step teaching plans, as well as plans for maintenance, self-direction, and generalization for each of the three goals. These goals were then added to the student's individualized education program (IEP), if not already present, in a separate meeting with the IEP team that occurred after the consultation. Following the initial consultation, the CT opened an envelope to reveal what type of follow-up would occur (two or four coaching sessions, e-feedback, no coaching). Teachers assigned to the consultation-only condition were contacted by the researchers at the end of the school year and interviewed about their student's progress on each of the goals in a procedure identical to that for the other conditions.

Coaching condition. Coaching sessions generally lasted 1 hour during which the CT, teacher, and caregiver, if able to

Table 1. Teacher, Student, and PET-GAS Characteristics by Dosage and Type Condition.

Characteristic	Dose 1	Dose 2	Dose 3	$F(df_1, df_2)$	p	η^2	Type 1	Type 2	Type 3	$F(df_1, df_2)$	p	η^2
Teachers ($n = 28$)												
n	10	9	9				10	12	6			
Gender												
Male	1	1	0				1	0	1			
Female	9	8	9				9	12	5			
Race/ethnicity												
White	10	9	9			.06	10	12	6			
Years of teaching experience				$F(2, 24) = 0.71$.50					$F(2, 25) = 0.50$.61	.04
M	8	13	12				8	12	14			
SD	5	9	10				5	10	7			
Years of autism teaching experience				$F(2, 17) = 1.29$.30	.13				$F(2, 19) = 0.19$.83	.02
M	8	11	4				8	9	11			
SD	5	9	3				6	8	8			
ASSET				$F(2, 24) = 0.12$.68	.03				$F(2, 24) = 0.78$.47	.06
M	3.21	3.22	3.42				3.21	3.43	3.11			
SD	0.67	0.52	0.48				0.67	0.50	0.44			
Students ($n = 28$)												
n	10	9	9				10	12	6			
Gender												
Male	9	5	9				9	9	5			
Female	1	4	0				1	3	1			
Race/ethnicity												
Black	0	0	2				0	1	1			
Multiracial	1	0	0				1	0	0			
White	9	9	7			.03	9	11	5			
SCQ-C				$F(2, 25) = 0.42$.66					$F(2, 25) = 0.60$.56	.05
M	23	22	20				23	20	23			
SD	9	6	4				9	5	4			

(continued)

Table 1. (continued)

Characteristic	Dose 1	Dose 2	Dose 3	$F(df_1, df_2)$	p	η^2	Type 1	Type 2	Type 3	$F(df_1, df_2)$	p	η^2
Vineland-TR												
M	56	71	52	$F(2, 16) = 0.51$.61	.06	56	64	47	$F(2, 16) = 2.73$.10	.25
SD	15	17	19				15	20	18			
Family annual household income												
<\$10,000	0	0	2	$F(2, 16) = 0.51$.61	.06	0	1	1	$F(2, 16) = 2.73$.10	.25
\$10,000–\$24,999	1	2	1				1	3	0			
\$25,000–\$49,999	4	2	2				4	3	1			
\$50,000–\$99,999	3	4	4				3	4	4			
\$100,000 or more	0	1	0				0	1	0			
PET-GAS ($n = 25$)												
n	7	9	9				7	12	6			
Difficulty												
M	2.85	2.79	2.78	$F(2, 21) = 0.16$.85	.14	2.86	2.76	2.83	$F(2, 21) = 0.30$.75	.03
SD	0.18	0.35	0.29				0.18	0.26	0.40			
Measurability												
M	2.57	2.67	2.78	$F(2, 22) = 0.61$.55	.24	2.57	2.72	2.72	$F(2, 22) = 0.41$.67	.04
SD	0.37	0.37	0.37				0.37	0.40	0.33			
Equidistance												
M	2.86	2.70	2.63	$F(2, 22) = 0.75$.49	.26	2.86	2.67	2.67	$F(2, 22) = 0.65$.53	.06
SD	0.26	0.39	0.42				0.26	0.40	0.42			

Note. The df_2 for baseline teacher and student variables vary due to missing data. Dose 1 = initial consultation only; Dose 2 = 1 coaching or emailed feedback; Dose 3 = 2 to 4 coaching or emailed feedback; Type 1 = initial consultation only; Type 2 = coaching; Type 3 = emailed feedback; PET-GAS = Psychometrically Equivalent Tested–Goal Attainment Scaling; ASSET = Autism Self-Efficacy Scale for Teachers (Ruble et al., 2013b, 1 “Not at all certain” to 4 “Very certain”, $\alpha = .95$); SCQ-C = Social Communication Questionnaire–Current (Rutter et al., 2003); TR = Vineland–Teacher Report (Sparrow et al., 2005).

attend, watched teacher-made videos of the implementation of the intervention plans for each goal. The videos were generally 3 to 5 minutes long and captured the entire teaching sequence described in the intervention plans. Based on these videos, the teacher and CT rated the student's goal-attainment progress using goal-attainment scaling (Ruble et al., 2012; described below), reviewed the intervention plans, problem-solved any challenges, and updated the plans to meet the student and teacher's changing needs. Caregivers and teachers received a copy of the coaching summary report compiled by the CT, which included a description of the video, the student's goal-attainment progress rating, and any changes made to the teaching plans following each coaching session. CTs then received supervision by the researchers for each coaching session based on an audio recording of the entire coaching session, the teacher's videos, and the coaching summary report (Ruble et al., 2022). Researchers independently rated observations of the student's goal-attainment progress and the teacher's adherence to the intervention plans via teacher-made videos of implementation for each goal.

E-feedback condition. E-feedback differed from coaching in that all communications with the teacher was via email. Teachers sent their CT any changes they made to the intervention plans along with videos of themselves implementing the step-by-step teaching plans for each goal. Based on the videos, the CT completed and emailed a PF report with their ratings of the student's goal-attainment progress and the teacher's percentage of adherence to the intervention plans for each goal. CTs also included comments on the report with suggestions for improving the teacher's implementation of the teaching plans (e.g., "Next time, remember to wait 5 seconds after prompting her to give her time to respond.") and offering encouragement (e.g., "Great job implementing these teaching plans! The student is really starting to make progress!").

Design

As mentioned, teachers were randomly assigned to receive one of four follow-up conditions at the end of the initial consultation by opening an envelope that revealed their condition. Those four conditions consisted of (a) four coaching sessions; (b) two coaching sessions, (c) four e-feedbacks, or (d) initial consultation only, no coaching or e-feedback. However, due to the COVID-19 pandemic and related school closures in the spring of 2020, participants were unable to complete the condition they were randomly assigned to. Four sets of teacher/caregiver/student participants despite being randomly assigned to receive either coaching or e-feedback were unable to complete any coaching sessions or e-feedback reports prior to schools closing to in-person instruction, while an additional 11 sets of

participants were unable to complete the full dosage of coaching or e-feedback they were assigned to receive. An additional three sets of participants had to be excluded from the study altogether because they were unable to schedule their initial consultations before schools closed and were unable to meet via video conferencing due to personal and technical challenges. As both coaching and e-feedback depend on videos of the teacher's implementation of the intervention plans with the student, neither could be done once schools closed to in-person instruction for the remainder of the 2020 academic year. Due to these challenges, data were recategorized for data analysis based on what participants were able to complete in terms of both dosage and type. These new categories of follow-up include three conditions for type: (a) initial consultation only, no e-feedback or coaching; (b) e-feedback; and (c) coaching. In addition, dosage conditions include (a) initial consultation only, no e-feedback or coaching; (b) one coaching/e-feedback; and (c) two to four coaching/e-feedback. Due to only four participants receiving four coaching/e-feedback sessions, this group was combined with a group of five additional participants who received two coaching/e-feedback sessions so that the three groups were relatively equal in number of participants.

Measures

Baseline measures to ensure sample equivalency. Several measures were used at baseline to ensure sample equivalency. All CTs, caregivers, and teachers completed baseline demographic surveys (see Table 1). Caregivers also completed a 40-item yes-no autism screener using the SCQ-C (Rutter et al., 2003) during recruitment. The SCQ-C evaluates communication skills and social functioning observed in the past 3 months in children who are suspected of having ASD, with higher scores representing more symptoms of ASD. Teachers also completed the Autism Self-Efficacy Scale for Teachers (ASSET; Ruble et al., 2013b), a 30-item self-report measure ($\alpha = .96$) that evaluates special education teacher's beliefs regarding their ability to perform tasks that are associated with teaching a specific student with ASD as well as the Vineland-II Adaptive Behavior Scales—Classroom Edition (Sparrow et al., 2005) to evaluate their student's personal independence and social skills used for everyday living.

Psychometrically equivalence tested goal attainment scaling. To assess student outcomes, goal-attainment scaling was used to rate the student's progress toward IEP goals. Psychometrically Equivalence Tested–Goal Attainment Scaling (PET-GAS) is an idiographic approach for measuring the outcomes of highly individualized goals and interventions in a way that nomothetic measures are not capable of doing (Ruble et al., 2022). First, student PET-GAS was

developed in collaboration with the CTs following the initial COMPASS consultation using a structured process to ensure psychometric equivalence (Ruble et al., 2012). Students' PET-GAS' were developed using a five-point scale ranging from of -2 to +2, with -2 representing the student's present level of performance at the time of the initial consultation, -1 representing progress toward the goal, 0 representing progress consistent with the stated annual goal, +1 representing progress exceeding the goal, and +2 representing progress greatly exceeding the goal (Ruble et al., 2013a, 2019, 2022).

Second, each PET-GAS form was formally measured for psychometric equivalency using a three-point ordinal scale (e.g., 1 "not at all," 2 "somewhat," 3 "very") to code three key components: (a) level of goal difficulty compared to the student's present levels of performance at the time of the consultation, (b) goal measurability (i.e., observable skill, prompt levels described, criterion for success described), and (c) equidistance of scaled objectives used to rate the goal (e.g., -1 description is 50% less than the goal) (Ruble et al., 2013a). Two raters independently coded 20% of the PET-GAS forms for psychometric equivalence. The sample interrater agreement was 0.89 for difficulty, 0.89 for measurability, and 0.55 for equidistance. Interrater agreement for equidistance between the goal statements was relatively lower than that for the other items (i.e., level of difficulty and measurability) due to some disagreement about whether the goals were equilibrated appropriately relative to the goal (e.g., prompt level, settings, frequency of skill increases/decreases by at least 50% for all descriptions at the -1, +1, and +2 levels). To ensure that the PET-GAS scales were psychometrically equivalent across dosage and follow-up type conditions, an analysis of variance was conducted and showed no statistically significant differences in mean scores of level of difficulty, measurability, and equidistance for the three dosages (initial consultation only, 1, and 2-4) and the three types of follow-up (coaching, feedback, and initial consultation only) (see bottom of Table 1).

Psychometrically Equivalence Tested-Goal Attainment Scaling progress was assessed at each coaching and emailed feedback session based on the direct observation of teacher-made videos. For the entire sample, including those who received no coaching or feedback following the initial consultation, progress was assessed at the end of the school year (i.e., final PET-GAS) by the researchers using a structured teacher interview and teacher progress monitoring data. To assess interrater agreement, two coders independently coded 20% of the PET-GAS scores, resulting in a sample intraclass correlation coefficient (ICC) for average measures based on absolute agreement of 0.96.

Teacher adherence to teaching plans. Teacher adherence to the teaching plans was assessed by the researchers following each coaching or emailed feedback session based on

observation from videos. Teaching plan adherence was rated using a 0 to 4 scale: 0 = 0% of components implemented, 1 = 1% to 25% implemented, 2 = 26% to 50% implemented, 3 = 51% to 75% implemented, 4 = 76% to 100% implemented. Adherence for the three goals was then averaged to obtain an overall mean score for each session. To assess interrater agreement, two coders independently rated 20% of the sessions, resulting in a sample interrater agreement of 0.92.

Analysis

First, conditions were compared based on key participant characteristics (Table 1) to ensure a similar baseline for all conditions. Given that student-teacher dyads are situated within consultants, our data have a two-level data structure. Given the small sample size of our study, random slopes were not feasible (estimable). Thus, only a random intercept for each outcome was included. Thus, we used two-level multilevel models (MLMs; Finch & Bolin, 2017) with random intercepts to evaluate the impact of dosage (initial consultation only, 1 coaching/e-feedback session, 2-4 coaching/e-feedback session) and type (initial consultation only, coaching, and e-feedback) on student PET-GAS scores with and without adjusting for time (days) in intervention. Finally, we used two-level MLMs with random intercepts to determine the impact of dosage (1 vs. 2-4) and type (coaching vs. e-feedback) on teachers' adherence to the intervention plans at the final coaching/e-feedback session with and without controlling for time (days) in intervention. To account for four cases with missing data on student PET-GAS scores, multilevel multiple imputation methods as implemented in Blimp 3 were used with 50 imputations (Keller & Enders, 2021). Given our interest was on fixed effects and not random effects, MLMs were conducted with maximum likelihood estimation with robust standard errors in Mplus 8.7 (2021). In addition, an MLM adjusting for time (days) in intervention was compared to an MLM without this covariate to identify the final model. This comparison was conducted by examining the difference in the two models adjusted for the Bayesian information criterion (BIC), with a smaller BIC value considered to have superior model fit. A BIC value difference exceeding 10 provides strong evidence of model fit difference (Kass & Raftery, 1995). Prior to model comparisons, a preliminary step was performed wherein an unconditional model (without predictors) was estimated to calculate the amount of variance in each outcome variable between consultants (ICC). All statistical significance tests were performed at the 5% significance level.

Considering the MLMs, the Level 1 model is situated at the student/teacher level, which included final PET-GAS or adherence scores. Using final PET-GAS and dose as an example, with days as a covariate, the Level 1 model was

$$\text{PET-GAS}_{ij} = \beta_{0j} + \beta_{1j}\text{DOSE_D1}_{ij} + \beta_{2j}\text{DOSE_D2}_{ij} + \beta_{3j}\text{DAYS}_{ij} + e_{ij}, \quad [1]$$

where PET-GAS_{ij} is the final PET-GAS score at the end of the intervention for student i with consultant j , DOSE_D1_{ij} is the dummy variable comparing initial consultation only to one coaching/feedback session, DOSE_D2_{ij} is the dummy variable comparing initial consultation only to two to four coaching/feedback sessions, DAYS is the number of days in intervention for the same student, and e_{ij} is an error term unique to each student, assuming $e_{ij} \sim N(0, \sigma^2)$. The average final PET-GAS score of students in the initial consultation-only condition is represented by β_{0j} for consultant j adjusted for time in study (days, grand mean centered) of students for that consultant; β_{1j} is the treatment effect of the one coaching/feedback session condition compared to the initial consultation-only condition for consultant j ; β_{2j} is the treatment effect of the two to four coaching/feedback sessions condition compared to the initial consultation-only condition for consultant j ; and β_{3j} is the regression coefficient for the number of days in intervention on PET-GAS score for consultant j .

The Level 2 model was situated at the consultant level, which did not include any covariates or predictors given that group assignment occurred at the student level. Therefore, the Level 2 model was

$$\beta_{0j} = g_{00} + u_{0j}, \beta_{0j} = g_{00} + u_{0j}, \quad [2]$$

$$\beta_{1j} = g_{10}, \quad [3]$$

$$\beta_{2j} = g_{20}, \quad [4]$$

$$\beta_{3j} = g_{30}, \quad [5]$$

where g_{00} is the average PET-GAS score for the initial consultation-only condition; g_{10} is the pooled regression coefficient for the treatment effect comparing one coaching/feedback condition to initial consultation-only condition; g_{20} is the pooled regression coefficient for the treatment effect comparing two to four coaching/feedback sessions condition to initial consultation-only condition; g_{30} is the pooled regression coefficient for days; and u_{0j} is an error term unique for each consultant, assuming $u_{0j} \sim N(0, \tau^2)$. The Level 2 model is empty and lacks an adjustment because group assignment was conducted at the student/teacher level. However, this model will account for the hierarchy of students nested within consultants.

Results

In total, 12 teachers received coaching, six received e-feedback, and 10 received the initial consultation only. Of the 12

who received coaching, five received one coaching session, five received two coaching sessions, and two received four coaching sessions. Of those who received e-feedback, four received one e-feedback report, and two received four e-feedback reports. These 28 cases were analyzed in two different ways by type (coaching vs. e-feedback vs. initial consultation only) and dose (initial consultation only, one, and two to four). The dosage categories include those who received either coaching or e-feedback. Those who received either two or four coaching sessions/e-feedback reports were combined into a single group for ease of analysis.

A summary of baseline characteristics of teacher and child factors is provided in Table 1. None of the teacher characteristics (i.e., years of teaching in general and with students with ASD in particular and self-efficacy for teaching students with ASD as measured by the ASSET; Ruble et al., 2013b) varied based on dosage and type of feedback. Furthermore, student characteristics at baseline (i.e., ASD severity and adaptive behavior ratings as measured by teacher reported Vineland-II; Sparrow et al., 2005) did not differ based on dosage and type. Finally, PET-GAS ratings were similar across conditions for level of goal difficulty, goal measurability, and equidistance of goal benchmarks (see Table 1).

Considering PET-GAS as an outcome, the average cluster size across nine consultants and 28 observations was 3.11 ($SD = .93$). When considering adherence, which was restricted to teachers who participated in an intervention condition where adherence could be measured (i.e., coaching and e-feedback), the average cluster size across eight consultants and 18 observations was 2.25 ($SD = .89$).

Does Dosage and Type of PF Matter for PET-GAS?

Table 2 provides the MLM results examining whether dosage of PF or type of PF influences student's PET-GAS scores adjusting or not adjusting for time (days) in intervention. Preliminary estimates for the unconditional model for PET-GAS indicated that 3.7% (i.e., $ICC = 0.05/[0.05 + 1.30]$) of the variance exists between consultants. A comparison of the adjusted BIC values for the unadjusted MLM (76.01) and adjusted MLM (74.19) for dosage shows a minimal difference between the models. A similar finding was observed for type (82.47 vs. 78.64). Therefore, results focus on reporting and interpreting the unadjusted MLMs.

Regarding dosage of PF, results indicate that students whose teachers received two to four coaching/feedback sessions had a higher observed PET-GAS mean ($M = 0.26$) than those whose teachers received the initial consultation only ($M = -1.00$), estimate = 1.26, $p = .02$, $ES = 1.27$, or one coaching/feedback session ($M = -0.81$), estimate = 1.08, $p < .001$, $ES = 1.09$. There was no difference in students' PET-GAS means between teachers receiving the initial consultation only versus one coaching/feedback session,

Table 2. MLM Results Estimating Effects of Type of PF Condition and Dosage on PET-GAS Unadjusted and Adjusted for Days in Study.

		Unadjusted				Adjusted			
Level	Parameter	Est.	SE	p	ES	Est.	SE	p	ES
Dose									
Within (Level 1: student)	ICO vs. 1 Coach/EF	0.19	0.69	.79	0.19	0.11	0.58	.85	0.11
	ICO vs. 2–4 Coach/EF	1.26	0.55	.02	1.27	1.03	0.59	.08	1.09
	1 vs. 2–4 Coach/EF	1.08	0.26	<.001	1.09	–0.92	0.4	.02	–0.97
	Days (centered)					0.01	0.01	.57	
	Residual variance	0.97	0.74	.19		0.88	0.64	.17	
Between (Level 2: consultant)	Intercept (mean)	–1.00	0.55	.07		–0.87	0.54	.1	
	Variance	0.01	0.4	.02		0.01	0.21	.95	
	Mean					Mean			
	ICO (<i>n</i> = 10)	–1.00				–0.87			
	1 Coach/EF (<i>n</i> = 9)	–0.81				–0.77			
	2–4 Coach/EF (<i>n</i> = 9)	0.26				0.15			
Type									
Within (Level 1: student)	ICO vs. Coach	0.55	0.68	.42	0.50	0.45	0.67	.5	0.44
	ICO vs. EF	0.63	0.72	.38	0.57	0.48	0.8	.54	0.48
	Coach vs. EF	0.08	0.53	.87	0.08	0.05	0.52	.93	0.05
	Days (centered)					0.01	0.01	.38	
	Residual variance	1.16	0.87	.18		0.98	0.68	.15	
Between (Level 2: consultant)	Intercept (mean)	–0.86	0.65	.18		–0.79	0.66	.23	
	Variance	0.06	0.34	.87		0.04	0.50	.94	
	Mean					Mean			
	ICO (<i>n</i> = 10)	–0.86				–0.79			
	Coach (<i>n</i> = 12)	–0.31				–0.34			
	EF (<i>n</i> = 6)	–0.23				–0.30			

Note. Results based on multilevel multiple imputation (*k*=50 imputations). MLM = multilevel model; PF = performance feedback; ICO = initial consultation only; Coach = coaching; EF = emailed feedback; Est. = coefficient estimate; ES = effect size ($\text{Est.}/\sqrt{(\text{residual variance} + \text{between variance})}$); PET-GAS = Psychometrically Equivalence Tested–Goal Attainment Scaling.

estimate = 0.19, *p* = .79, ES = 0.19. However, no significant differences were observed for type of PF on PET-GAS (see Table 2).

Does Dosage and Type of PF Matter for Teacher Adherence?

Table 3 provides the MLM results examining whether dosage or type of PF influences teacher adherence adjusting or not adjusting for time (days) in intervention. Preliminary estimates for the unconditional model for teacher adherence indicated that 0.01% (i.e., ICC = 0.01/[0.01 + 1.05]) of the variance exists between consultants. A comparison of the adjusted BIC values for the unadjusted MLM (46.60) and adjusted MLM (41.61) for dosage shows a minimal difference between the models. A similar finding was observed for type (50.03 vs. 48.63). Therefore, results focus on reporting and interpreting the unadjusted MLMs.

Similar to the findings of student PET-GAS, there was a significant effect for dose, estimate = –1.43, *p* = .001, ES = 1.43, with the teachers who received two to four sessions (*M* = 3.56) attaining higher adherence at their last session than those who only received one session (*M* = 2.13). However, there was no significant effect for type of PF, estimate = 0.57, *p* = .21, ES = 0.58.

Discussion

Research on the amount and type of teacher coaching is important because little is understood about how much coaching with PF is sufficient and whether face-to-face coaching vs. e-feedback matters. Analysis of baseline measures that could explain outcomes revealed that teacher factors, such as teaching experience and self-efficacy (ASSET; Ruble et al., 2013b), and child factors, such as ASD symptoms (SCQ-C; Rutter et al., 2003), adaptive

Table 3. MLM Results Estimating Effects of Type of PF Condition and Dosage on Teacher Adherence Unadjusted and Adjusted for Days in Study.

Level	Parameter	Unadjusted				Adjusted			
		Est.	SE	p	ES	Est.	SE	p	ES
Dose									
Within (Level 1: teacher)	1 vs. 2–4 Coach/EF Days (centered)	−1.43	0.43	.001	1.43	−1.44	0.36	<.001	1.81
	Residual variance	0.47	0.26	.07		0.53	0.23	.02	
Between (Level 2: consultant)	Intercept (mean)	3.56	0.30	<.001		3.49	0.27	<.001	
	Variance	0.53	0.43	.22		0.10	0.28	.72	
		Mean				Mean			
	1 Coach/EF (<i>n</i> = 9)	2.13				2.05			
	2–4 Coach/EF (<i>n</i> = 9)	3.56				3.49			
Type									
Within (Level 1: teacher)	Coach vs. EF Days (centered)	0.57	0.46	.21	0.58	0.52	0.44	.23	0.54
	Residual variance	0.96	0.22	<.001		0.91	0.21	<.001	
Between (Level 2: consultant)	Intercept (mean)	2.39	0.42	<.001		2.42	0.33	<.001	
	Variance	0.02	0.23	.92		0.01	0.31	.99	
		Mean				Mean			
	Coach (<i>n</i> = 12)	2.96				2.94			
	EF (<i>n</i> = 6)	2.39				2.42			

Note. MLM = multilevel model; PF = performance feedback; Coach = coaching; EF = emailed feedback; Est. = coefficient estimate; ES = effect size (Est./ $\sqrt{(\text{residual variance} + \text{between variance})}$).

behavior (Vineland; Sparrow et al., 2005), and PET-GAS quality (Ruble et al., 2012), were unrelated to dosage and type of PF.

Our findings suggest that teacher adherence and student goal-attainment outcomes depend on dosage, not type of coaching. Specifically, having two or four sessions produced stronger results than having the initial consultation only or only one PF session. This finding is concordant with prior research that consultation with consistent follow-up coaching is essential to see improved teacher implementation and student outcomes (Kretlow & Bartholomew, 2010; Noell et al., 2014; Sleiman et al., 2020). However, it contrasts with the meta-analytic findings of Kraft et al. (2018) that coaching dosage (both total hours of one-to-one coaching and total hours of PD when coaching is paired with other PD) was unrelated to either teacher's instructional effectiveness or student outcomes. Using Kraft et al.'s (2018) dosage criteria, however, standard COMPASS with 4 hours of coaching or 7 hours of total PD including the initial consultation would be classified among the lowest-dosage coaching interventions (10 or less hours) included in their meta-analysis.

Their finding that coaching quality may be more important than coaching quantity is consistent with the argument that both COMPASS coaching and e-feedback with PF are high-quality interventions that incorporate many of the best practices in teacher PD (Kraft et al., 2018; Sleiman et al.,

2020). The Collaborative Model for Promoting Competence and Success PF, in keeping with Kraft et al.'s (2018) recommendations for high-quality coaching, is individualized, intensive, sustained, context-specific, and focused on specific skills related to high-quality teaching practices. In addition, we expanded our evaluation of COMPASS to include a multidimensional assessment of implementation that includes measures of acceptability, feasibility, appropriateness, and fidelity (i.e., implementation adherence and quality of delivery) that is explored in Ruble et al. (2022). We found COMPASS to be highly acceptable, feasible, and appropriate from the perspective of CTs, teachers, and caregivers and that novice CTs could be trained to implement COMPASS with high fidelity when provided with appropriate training and supervision by the researchers (Ruble et al., 2022).

It was surprising that both e-feedback and face-to-face coaching produced similar outcomes following the initial consultation. A critical commonality was that both types of follow-up provided teacher PF on intervention plan adherence and PET-GAS student outcomes based on video analysis of teacher-student interactions, which research has repeatedly noted as critical to the successful implementation of interventions (Fallon et al., 2015; Noell et al., 2014; Sleiman et al., 2020). However, there were important differences between the two types of follow-up.

For teachers in the coaching condition, they were permitted to engage in interactive problem-solving with guided questioning from the COMPASS consultant and opportunity for self-reflection and verbal praise when they or their student performed well. Teachers in the e-feedback condition also received written suggestions, encouragement, and praise along with the student's PET-GAS scores and teacher's adherence scores. The inclusion of these suggestions and praise may have strengthened the effectiveness of the emailed feedback condition to make it more comparable to face-to-face coaching. Sleiman et al. (2020) found that the inclusion of largely positive and encouraging comments as opposed to providing only constructive criticism was associated with higher effect scores overall, and our findings are consistent with this observation. Dismantling these components may be important for future research.

Limitations, Future Directions, and Implications

Our study has several limitations. First, we realize that from a design perspective, our small number of clusters (i.e., nine clusters for PET-GAS and eight clusters for adherence) is not advised for MLMs (McNeish & Stapleton, 2016). However, despite being small, our sample size is comparable to that of other studies investigating novel interventions and other studies of consulting interventions (Wainer et al., 2017). We also acknowledge that even though our models converged with our small sample size, some estimates may contain bias, which may affect inferences being made. However, as discussed in McNeish and Stapleton's (2016) review of simulation studies on the effect of small sample size on two-level models estimated with maximum likelihood, the number of clusters in our study does meet the recommended minimum number of clusters (at least five) to accurately estimate Level 1 fixed effects. As such, our point estimates for group differences can be treated as unbiased. However, the Level 1 fixed-effect standard errors may be downwardly biased (underestimated) as they were estimated well below the minimum recommended number of clusters (at least 30) for accurate estimation. Thus, our inferences about group differences may be viewed as false positives; however, this is purely speculation until this study is replicated with a larger number of clusters. In addition, our Level 1 and Level 2 variance estimates may not be accurate as they were estimated with fewer than the minimum recommended number of clusters of 10 and 30, respectively. However, point estimates for the Level 1 variance are minimally affected by sample size at either level. Furthermore, our Level 2 variance estimates may be deflated given the small number of clusters and small number of observations per cluster. Despite these limitations of the analytic model, our primary purpose was to develop and test a training program in COMPASS for school-based CTs and provide preliminary estimates for the fixed treatment

effect for dose and PF (not random effects) which could be used as point estimates for future planned data analyses. Therefore, a smaller sample was useful as we refined our training package using an iterative approach. Finally, no a priori power analysis was conducted prior to the performed data analyses as we did not expect to realize a power of 0.80 with large ES given the iterative nature of our approach. Additional research is needed with a larger, more diverse sample of CTs, teachers, caregivers, and students.

Unfortunately, due to the onset of COVID-19, we were unable to complete the study as planned, which further reduced our sample sizes as CTs were unable to conduct initial consultations and coaching/e-feedback sessions with teachers and caregivers once schools were closed to in-person instruction in the spring of 2020. Future research should investigate the impact of dosage and type of PF based on the original RCT we designed (four coaching, two coaching, four e-feedback, and initial consultation only with no follow-up coaching or e-feedback).

Another limitation of our study is that PET-GAS and adherence were treated as continuous outcomes during the data analyses, but they are not truly equidistance measures. However, given the small sample size, we were unable to treat the constructs that these measures represent as continuous in nature, which could be made possible within a latent variable framework such as multilevel structural equation modeling or multilevel item response modeling. As such, we leave this avenue for future research to explore.

Despite these limitations, our findings are promising that e-feedback may be a viable alternative or complement to coaching and is worthy of further examination in future research. Overall, these findings suggest that providing teachers more than one opportunity to receive PF via coaching or e-feedback following the development of high-quality intervention plans is essential for increasing their adherence to the plans and increasing student outcomes. However, more research on the frequency and number of coaching/e-feedback sessions is needed as schools plan for utilizing the valuable time of coaches and consultants.

Authors' Note

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Mental Health or the National Institutes of Health.

Acknowledgments

We thank school administrators who allowed their consultants and teachers the extra time and effort to help us learn about COMPASS and how to best support teachers. We are also grateful to parents and students who generously gave their time to help us learn. We would also like to thank Abigail Love for helping with the initial training package and Kahyah Pinkman for her assistance with data entry and coding.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This work was supported by the grant R34MH111783 from the National Institute of Mental Health.

ORCID iDs

Lindsey N. Ogle  <https://orcid.org/0000-0002-9261-4055>

Lisa A. Ruble  <https://orcid.org/0000-0001-7495-6572>

John H. McGrew  <https://orcid.org/0000-0003-2874-3896>

References

- Beidas, R. S., Edmunds, J. M., Marcus, S. C., & Kendall, P. C. (2012). Training and consultation to promote implementation of an empirically supported treatment: A randomized trial. *Psychiatric Services*, 63(7), 660–665. <https://doi.org/10.1176/appi.ps.201100401>
- Brock, M. E., & Carter, E. W. (2017). A meta-analysis of educator training to improve implementation of interventions for students with disabilities. *Remedial and Special Education*, 38(3), 131–144. <https://doi.org/10.1177%2F0741932516653477>
- Brock, M. E., Dynia, J. M., Dueker, S. A., & Barczak, M. A. (2020). Teacher-reported priorities and practices for students with autism: Characterizing the research-to-practice gap. *Focus on Autism and Other Developmental Disabilities*, 35(2), 67–78. <https://doi.org/10.1177/1088357619881217>
- Dunst, C. J., Bruder, M. B., & Hamby, D. W. (2015). Meta-synthesis of in-service professional development research: Features associated with positive educator and student outcomes. *Educational Research and Reviews*, 10, 1731–1744. <https://doi.org/10.5897/ERR2015.2306>
- Ennis, R. P., Royer, D. J., Lane, K. L., & Dunlap, K. D. (2020). The impact of coaching on teacher-delivered behavior-specific praise in pre-K–12 settings: A systematic review. *Behavioral Disorders*, 45, 148–166. <https://doi.org/10.1177%2F0198742919839221>
- Fallon, L. M., Collier-Meek, M. A., Maggin, D. M., Sanetti, L. M., & Johnson, A. H. (2015). Is performance feedback for educators an evidence-based practice? A systematic review and evaluation based on single-case research. *Exceptional Children*, 81, 227–246. <https://doi.org/10.1177%2F0014402914551738>
- Finch, W. H., & Bolin, J. E. (2017). *Multilevel modeling using Mplus*. CRC Press.
- Fraser, D. W., Marder, T. J., deBettencourt, L. U., Myers, L. A., Kalyon, K. M., & Harrell, R. M. (2020). Using a mixed-reality environment to train special educators working with students with autism spectrum disorder to implement discrete trial teaching. *Focus on Autism and Other Developmental Disabilities*, 35, 3–14. <https://doi.org/10.1177/1088357619844696>
- Hamrick, J., Cerda, M., O'Toole, C., & Hagen-Collins, K. (2021). Educator knowledge and preparedness for educating students with autism in public schools. *Focus on Autism and Other Developmental Disabilities*, 36, 213–224. <https://doi.org/10.1177/1088357621989310>
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90, 773–795. <https://doi.org/10.1080/01621459.1995.10476572>
- Keller, B. T., & Enders, C. K. (2021). *Blimp user's guide* (Version 3). www.appliedmissingdata.com/multilevel-imputation.html
- Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research*, 88(4), 547–588. <https://doi.org/10.3102%2F0034654318759268>
- Kretlow, A. G., & Bartholomew, C. C. (2010). Using coaching to improve the fidelity of evidence-based practices: A review of studies. *Teacher Education and Special Education*, 33(4), 279–299. <https://doi.org/10.1177%2F0888406410371643>
- Mahoney, M. (2020). Implementing evidence-based practices within multi-tiered systems of support to promote inclusive secondary classroom settings. *Journal of Special Education Apprenticeship*, 9, Article 2. <https://scholarworks.lib.csusb.edu/josea/vol9/iss1/2>
- McLeskey, J., Barringer, M.-D., Billingsley, B., Brownell, M., Jackson, D., Kennedy, M., Lewis, T., Maheady, L., Rodriguez, J., Scheeler, M. C., Winn, J., & Ziegler, D. (2017, January). *High-leverage practices in special education*. Council for Exceptional Children & CEEDAR Center. <http://ceedar.education.ufl.edu/wp-content/uploads/2017/07/CEC-HLP-Web.pdf>
- McNeish, D. M., & Stapleton, L. M. (2016). The effect of small sample size on two-level model estimates: A review and illustration. *Educational Psychology Review*, 28, 295–314. <https://doi.org/10.1007/s10648-014-9287-x>
- Morin, K. L., Ganz, J. B., Vannest, K. J., Haas, A. N., Nagro, S. A., Peltier, C. J., Fuller, M. C., & Ura, S. K. (2019). A systematic review of single-case research on video analysis as professional development for special educators. *The Journal of Special Education*, 53, 3–14. <https://doi.org/10.1177%2F0022466918798361>
- Morrier, M. J., Hess, K. L., & Heflin, L. J. (2011). Teacher training for implementation of teaching strategies for students with autism spectrum disorders. *Teacher Education and Special Education*, 34, 119–132. <https://doi.org/10.1177%2F0888406410376660>
- MPLUS. (2021) (Version 8.7) [Computer software]. Muthén & Muthén.
- Nagro, S. A., & Cornelius, K. E. (2013). Evaluating the evidence base of video analysis: A special education teacher development tool. *Teacher Education and Special Education*, 36, 312–329. <https://doi.org/10.1177%2F0888406413501090>
- National Research Council. (2001). *Educating children with autism*. National Academies Press. <https://doi.org/10.17226/10017>
- Noell, G. H., Gansle, K. A., Mevers, J. L., Knox, R. M., Mintz, J. C., & Dahir, A. (2014). Improving treatment plan implementation in schools: A meta-analysis of single subject design studies. *Journal of Behavioral Education*, 23, 168–191. <https://doi.org/10.1007/s10864-013-9177-1>
- Ruble, L. A., Dalrymple, N. J., & McGrew, J. H. (2010). The effects of consultation on individualized education program outcomes for young children with autism: The collaborative model for promoting competence and success. *Journal of Early Intervention*, 32, 286–301. <https://doi.org/10.1177%2F1053815110382973>

- Ruble, L. A., Dalrymple, N. J., & McGrew, J. H. (2012). *Collaborative model for promoting competence and success for students with ASD*. Springer.
- Ruble, L. A., Love, A. M. A., Wong, V. W., Grisham-Brown, J. L., & McGrew, J. H. (2020). Implementation Fidelity and Common Elements of High Quality Teaching Sequences for Students with Autism Spectrum Disorder in COMPASS. *Research in Autism Spectrum Disorders*, 71, <https://doi.org/10.1016/j.rasd.2019.101493>
- Ruble, L. A., & McGrew, J. H. (2015). *COMPASS and implementation science: Improving educational outcomes of children with ASD*. Springer.
- Ruble, L. A., McGrew, J. H., Dale, B., & Yee, M. (2021). Goal attainment scaling: An idiographic measure sensitive to parent and teacher report of IEP goal outcome assessment for students with ASD. *Journal of Autism and Developmental Disorders*, 52, 3344–3352. <https://doi.org/10.1007/s10803-021-05213-3>
- Ruble, L. A., McGrew, J. H., Toland, M. D., Dalrymple, N. J., Adams, M., & Snell-Rood, C. (2018). Randomized control trial of COMPASS for improving transition outcomes of students with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 48, 3586–3595. <https://doi.org/10.1007/s10803-018-3623-9>
- Ruble, L. A., McGrew, J. H., Toland, M. D., Dalrymple, N. J., & Jung, L. A. (2013a). A randomized controlled trial of COMPASS web-based and face-to-face teacher coaching in autism. *Journal of Consulting and Clinical Psychology*, 81, 566–572. <https://psycnet.apa.org/doi/10.1037/a0032003>
- Ruble, L. A., Ogle, L. N., & McGrew, J. (2022). Practice makes proficient: Evaluation of implementation fidelity following COMPASS consultation training. *Psychology in the Schools*, 60(3), 743–760. <https://doi.org/10.1002/pits.22800>
- Ruble, L. A., Toland, M., Birdwistell, J., McGrew, J. H., & Usher, E. (2013b). Preliminary study of the autism self-efficacy scale for teachers (ASSET). *Research in Autism Spectrum Disorders*, 7, 1151–1159. <https://doi.org/10.1016/j.rasd.2013.06.006>
- Ruble, L. A., Wong, W., & McGrew, J. H. (2019). Data collection in education and measurement of progress. In R. Jordan, K. Hume, & J. Roberts (Eds.), *The SAGE handbook of autism and education* (1st ed., pp. 578–592). SAGE.
- Rutter, M., Bailey, A., & Lord, C. (2003). *The social communication questionnaire: Manual*. Western Psychological Services.
- Sam, A. M., Odom, S. L., Tomaszewski, B., Perkins, Y., & Cox, A. W. (2021). Employing evidence-based practices for children with autism in elementary schools. *Journal of Autism and Developmental Disorders*, 51, 2308–2323. <https://doi.org/10.1007/s10803-020-04706-x>
- Sleiman, A. A., Sigurjonsdottir, S., Elnes, A., Gage, N. A., & Gravina, N. E. (2020). A quantitative review of performance feedback in organizational settings (1998–2018). *Journal of Organizational Behavior Management*, 40, 303–332. <https://doi.org/10.1080/01608061.2020.1823300>
- Sparrow, S. S., Balla, D. A., & Cicchetti, D. V. (2005). *Vineland adaptive behavior scales: Survey forms manual*. AGS Publishing.
- Wainer, A. L., Pickard, K., & Ingersoll, B. R. (2017). Using web-based instruction, brief workshops, and remote consultation to teach community-based providers a parent-mediated intervention. *Journal of Child and Family Studies*, 26, 1592–1602. <https://doi.org/10.1007/s10826-017-0671-2>
- Wong, V., Ruble, L. A., McGrew, J. H., & Yu, Y. (2018). An empirical study of multidimensional fidelity of COMPASS consultation. *School Psychology Quarterly*, 33, 251–263. <https://psycnet.apa.org/doi/10.1037/spq0000217>