


---

# Using Coaching to Improve the Fidelity of Evidence-Based Practices: A Review of Studies

Teacher Education and Special Education  
33(4) 279–299  
© 2010 Teacher Education Division of  
the Council for Exceptional Children  
Reprints and permission:  
sagepub.com/journalsPermissions.nav  
DOI: 10.1177/0888406410371643  
http://tese.sagepub.com  


Allison Graves Kretlow<sup>1</sup>  
and Christina C. Bartholomew<sup>1</sup>

## Abstract

The authors conducted a comprehensive review of research to identify the impact of coaching on changes in preservice and in-service teachers' implementation of evidence-based practices. They identified a total of 13 studies from the 20 years of literature they searched. In general, coaching improved the extent to which teachers accurately implement evidence-based practices such as ClassWide Peer Tutoring, Direct Instruction, Learning Strategies, and Positive Behavior Support in classrooms or practicum settings. The retrieved studies also suggest that highly engaged, small-group initial training, followed by multiple observations, feedback, and modeling are critical components across coaching interventions. A few studies also provide promising data to support the consequential effects of coaching on improvements in student achievement. The authors offer suggestions for future research and practice related to preservice and in-service teacher training.

## Keywords

evidence-based practices, meta-analysis, professional development, teacher preparation practices and outcomes

Increasing the use of evidence-based practices and improving the fidelity with which teachers implement them is a critical variable for maximizing student achievement. The research-to-practice gap is well documented (Abbott, Walton, Tapia, & Greenwood, 1999; Bulgren et al., 2002), as are typically low levels of fidelity (Scheeler, Bruno, Grubb, & Seavey, 2009; Swanson & Hoskyn, 1998). Although there are few studies in the education literature that specifically examine levels of fidelity, the ones that exist show that without substantial, ongoing support, fidelity levels are likely to stay low (DiGennaro, Martens, & Kleinmann, 2007).

Some research has shown a strong link between high fidelity and improved academic

achievement. For example, Furtak and colleagues (2008) and Kovalesski, Gickling, Morrow, and Swank (1999) showed that optimal gains in student achievement can be diminished when teaching procedures are inconsistently used. In other words, low or inconsistent levels of fidelity with teaching procedures correlated with lower gains in student achievement. Other data show it takes a great deal of effort to help teachers reach high

---

<sup>1</sup>James Madison University, Harrisonburg, Virginia, USA

## Corresponding Author:

Allison Graves Kretlow, James Madison University,  
Department of Exceptional Education, MSC 6908, 395  
S. High St., Harrisonburg, VA 22807, USA  
Email: kretloag@jmu.edu

fidelity. Specifically, Buzhardt, Greenwood, Abbott, and Tapia (2007) reported that improving fidelity can take years of intensive work with schools and individual teachers but that when this kind of ongoing support was provided, fidelity levels doubled. Similarly, Klingner, Vaughn, Hughes, and Arguelles (1999) found that teachers who received more ongoing support had higher fidelity and higher student achievement in their classrooms.

Preservice and in-service teacher preparation has the potential to be one of the most influential factors in increasing teachers' fidelity of evidence-based practices (McLeskey & Billingsley, 2008); however, both groups of teachers face similar barriers to the implementation of evidence-based practices in actual classroom settings. For preservice teachers, instructors often address evidence-based practices in courses but provide little follow-up support in practicum or student teaching settings (Scheeler et al., 2009). Similarly, for in-service teachers, the most common form of training is the one-day in-service, with limited follow-up support. However, professional development literature overwhelmingly shows this model leads to little change in teaching practice (Yoon, Duncan, Lee, Scarloss, & Shapley, 2007). In order for teachers to successfully implement any new practice, at some point the focus of training must account for the challenges encountered when teaching real students in real classrooms (Elmore, 2006).

Teachers face many challenges when beginning to implement any new practice. For example, teachers in one study reported that implementing new practices after professional development was difficult because of "not having an in-depth understanding of the practice," "forgetting how to use it correctly," or "needing a refresher" due to the complexity of the practice among many other classroom responsibilities (Klingner et al., 1999, p. 271). In a similar study, teachers reported receiving little training on the programs and practices they were expected to use, and therefore they implemented only the components that "worked for them" (Bartholomew & Kretlow, 2010). In addition, initial group

training sessions may not be a sufficient tool to maximize fidelity. Kretlow, Wood, and Cooke (2009) and Kretlow, Cooke, and Wood (in press) found that although the frequency with which teachers used evidence-based strategies increased after an initial training, high and stable levels of accurate implementation did not occur until after teachers received at least one individualized coaching session.

One way of improving fidelity is providing teachers with individualized follow-up support after initial training. One particularly effective form of follow-up support is coaching (Filcheck, McNeil, Greco, & Bernard, 2004; Jager, Reezigt, & Creemers, 2002; Kohler, Crilley, Shearer, & Good, 1997; Kohler, Ezell, & Paluselli, 1999; Kretlow et al., 2009, in press; Lignuaris-Kraft & Marchand-Martella, 1993; Maheady, Harper, Mallette, & Karnes, 2004; Miller, Harris, & Watanabe, 1991; Morgan, Menlove, Salzberg, & Hudson, 1994; Peck, Killen, & Baumgart, 1989; Pierce & Miller, 1994; Stitcher, Lewis, Richter, Johnson, & Bradley, 2006). Coaching involves an expert (e.g., university faculty or supervisor, lead teacher, skilled peer) providing individualized support to teachers after an initial training occurs. The purposes of coaching are to encourage accurate and sustained implementation of new teaching behaviors and to prevent the isolation that often occurs after teachers begin to implement new practices in their classrooms (Joyce & Showers, 1995). Coaching is intended to provide teachers with a "means of examining and reflecting on what they do in a psychologically safe environment where it is all right to experiment, fail, revise, and try again" (Raney & Robbins, 1989, p. 37).

During coaching, teachers are prompted to use a newly learned practice in a classroom setting while receiving feedback on their performance from an expert. From a behavioral standpoint, this model of training may make teachers more likely to use the techniques again because the new teaching behaviors are directly reinforced in the setting where instruction typically occurs (Scheeler et al., 2009).

For example, if a teacher tries a new practice and receives positive feedback from a coach (e.g., “Yes, that strategy worked perfectly because the students were highly engaged and responding with correct answers!”), the probability of the teacher using the technique accurately will likely go up. In contrast, if a teacher tries a new practice but makes some errors, the coach might model the strategy correctly and then prompt the teacher to try it again. Once the teacher uses the strategy correctly, the coach would give praise—again making the teacher more likely to use the strategy correctly in the future. In the absence of specific feedback, the reinforcers for a teacher are probably very different. For example, teachers may not be aware they are implementing a strategy incorrectly and may continue doing it because they or their students like it, or teachers may not have success with a strategy and discontinue it because they do not feel it works or do not have access to help.

Professional development literature suggests two dominant models of coaching: supervisory coaching (Joyce & Showers, 1995) and side-by-side (i.e., *in vivo*) coaching (Blakely, 2001). Peer coaching is also a subcategory of coaching that can be supervisory or side by side (Allen & LeBlanc, 2004). Using the supervisory follow-up method, the coach conducts an observation of a teacher implementing a technique he or she has recently learned to use in a prior training. During the observation, the coach records the presence or absence of particular instructional techniques the teacher was instructed to use in the initial training. After the lesson, the coach provides descriptive, nonevaluative feedback to the teacher regarding the strengths and opportunities for improvement noticed during the observation. Researchers have found that when supervisory follow-up coaching is provided in combination with an initial professional development opportunity, teaching accuracy improves (Fuchs, Fuchs, Hamlett, & Ferguson, 1992; Kohler et al., 1999).

The second model of coaching, the side-by-side coaching method, allows teachers to receive *in vivo* feedback specific to the

accuracy of their implementation of new teaching behaviors. In addition, side-by-side coaching allows teachers an opportunity to observe specific teaching procedures demonstrated by an expert (i.e., the coach) with their own students in the context of a real classroom lesson. During a side-by-side coaching session, the coach directly intervenes during the lesson, provides a model and a rationale for the change, and then provides additional opportunities for the teacher to teach the same format again with immediate feedback from the coach. Experimental investigations have shown that side-by-side coaching can improve the rate of acquisition of new teaching behaviors (Kohler et al., 1997) and the accuracy of teaching behaviors (Kretlow et al., 2009, *in press*) and can result in longer maintenance of accurate teaching behaviors than the supervisory follow-up method (O'Reilly & Renzaglia, 1992).

Another factor that makes coaching an effective model of follow-up support is that it integrates components of effective adult learning. Although there continues to be debate in the literature related to adult learning theories, Gordon (2004) suggests there is substantial consensus on key principles related to adult learning (i.e., Caffarella & Barnett, 1994; Merriam, 2001). Adults engage in the learning process when the content and/or new knowledge relate to their current experiences and they are allowed to actively participate in the learning process. “Adult learners are both autonomous and collaborative” (Gordon, 2004, p. 20); thus, coaching as a form of professional development allows teachers to practice their own skills in collaboration with a coach/consultant. Furthermore, because adults bring considerable life experiences to the learning process, it is important to integrate time for reflection during professional development. A coaching model supports these principles by allowing teachers to explore the application of new concepts within the context of their classroom and by allowing time for reflection and feedback.

Adult learners enter professional development at different points in their careers and

with varying degrees of knowledge and skills. Coaching allows professional development to address the individual teacher's needs related to her or his implementation of new practices. Furthermore, it moves professional development beyond abstract theories and educational principles and addresses authentic everyday challenges faced by teachers.

Research suggests that effective professional development for teachers includes sustained activities that allow teachers to apply the new content to their planning and instruction (Darling-Hammond, Wei, Andree, Richardson, & Orphanos, 2009; Desimone, Porter, Garet, Yoon, & Birman, 2002). Given the importance of fidelity in improving student achievement, it is critical to examine specific models of improving teaching accuracy that have resulted in observable changes in teachers' classroom practices, such as coaching. Therefore, the purpose of this review is to summarize studies that examined the impact of coaching interventions on observable changes in teachers' fidelity of evidence-based practices in classroom settings.

## Method

### Literature Search Procedures

**Electronic search.** Electronic searches were conducted using Academic Search Premier, ERIC, MasterFILE Premier, PsychINFO, PsychArticles, and Education Research Complete from 1989 to 2009. When searching electronic databases, the following keyword combinations were used: coaching; coaching and in-service; coaching and preservice; side-by-side; coaching and teach\*; peer coaching; supervisory coaching; coaching and reading; coaching and writing; coaching and spelling; coaching and behavior; coaching and positive behavior support; coaching and math; coaching and evidence-based; and coaching and research-based. The following journals were hand searched to locate the most recent studies (2008–2009): *Exceptional Children*, *Education & Treatment of Children*, *Journal of Professional Development*, *Journal of Staff*

*Development*, *Journal of Special Education*, *Remedial and Special Education*, and *Teacher Education and Special Education*. Websites of the journals listed above were also searched for online first articles—forthcoming articles published ahead of print. Finally, references from relevant studies were examined to locate additional articles. Unpublished dissertations were not included in the search.

**Selection of studies.** The purpose of this review was to summarize studies that examined the impact of a specific coaching intervention on quantitatively measured changes in teachers' classroom practices. Therefore, only studies that used a research design that allowed for causal inference were included (i.e., experimental, quasi-experimental, single-subject). Studies meeting the additional following criteria were included in the review: (a) it was published in a peer-reviewed journal; (b) participants were preservice or in-service teachers in general or special education working with students from preK to Grade 12; (c) the independent variable was coaching (i.e., supervisory or side-by-side); (d) the dependent variable was a direct, observable measure of specific instructional characteristics; and (e) coaching was related to an evidence-based practice with support for improving academic performance or appropriate classroom behavior.

Following the framework provided by Burns and Ysseldyke (2009) in their survey of reported evidence-based practices in special education, only practices with evidence of effectiveness derived from meta-analytic research were included. Once a study fit criteria (a) through (d) above, the authors searched for meta-analytic research on the practice coached in the study. Only practices with moderate to high effect sizes (i.e., greater than .20 weighted effect size) were included in this review. These practices included: Active Student Responding (Randolph, 2007); Positive Behavior Support (Office of Special Education Programs, 2010); ClassWide Peer Tutoring (Institute of Education Sciences, 2007); Collaborative Strategic Reading (Forness, 2001); direct

instruction/Direct Instruction (Promising Practices Network, 2005); and Learning Strategies using the Strategic Intervention Model (Kavale & Forness, 1999). Practices with small effects (e.g., modality training, social skills training programs) were excluded.

### Reliability

We initially retrieved 457 studies through electronic and hand searches. After initial retrieval, we calculated two types of reliability. Because of the large yield, 20% of the search terms were randomly selected for reliability. First, the reliability of article retrieval was calculated. The first author and a graduate assistant in special education independently searched the same randomly selected terms. The first author used a point-by-point method to calculate the number of agreements divided by the number of agreements plus disagreements, resulting in a reliability coefficient of 98.2%.

Second, the reliability of article inclusion was calculated. The second author independently reviewed a randomly selected 20% of the articles retrieved, applied the selection criteria, and marked the article for inclusion or exclusion in the review. The first author used the point-by-point method by dividing the number of agreements by the number of agreements plus disagreements, resulting in an inclusion reliability coefficient of 100%.

### Results

Of the 457 articles initially retrieved, 13 fit the selection criteria and were included in this review. See Table 1 for a detailed description of the individual studies. Because of the low number of studies meeting criteria (i.e., fewer than 20), results will be presented by number instead of percentage. Most articles were excluded because of research design (i.e., they were qualitative) or measurement of the dependent variable (i.e., not a direct measure of change, such as teacher self-report). Some studies were excluded because low effect

sizes had been documented in previous meta-analytic research for the practice coached (e.g., modality training, social skills programs). Ten studies used a single-subject, multiple-baseline-across-teachers design, and three studies used a group design (i.e., experimental, quasi-experimental). Seven studies were published before 2000, six studies were published after 2000, and the most recent study was published in 2010. Ten studies involved in-service teachers and three studies involved preservice teachers. Participants in seven studies were general educators; participants in the other six studies were special educators. One study involved preschool students, two studies involved middle/secondary students, and ten studies involved elementary students. The number of teacher participants for each study ranged from 1 to 32. Coaching was provided in math for two studies, reading for four studies, spelling for one study, multiple academic areas for four studies, and classroom management for two studies.

### Preservice and In-Service Teachers

In the 13 studies, a total of 110 teachers received coaching. Thirty-seven were in-service teachers (Filcheck et al., 2004; Kohler et al., 1997, 1999; Kretlow et al., 2009, in press; Miller et al., 1991; Peck et al., 1989; Stitcher et al., 2006) and 73 were preservice teachers (Hasbrouck, 1997; Lignugaris-Kraft & Marchand-Martella, 1993; Maheady et al., 2004; Morgan et al., 1994; Pierce & Miller, 1994). Forty-one were general education teachers, and 69 were special education teachers. All in-service teachers taught preschool or elementary students; therefore, none of the coaching interventions were delivered in a middle or high school setting. Teacher experience across studies ranged from 2 years to 30 years, although not all studies reported years of teacher experience. In most cases, teachers volunteered to participate in coaching activities. However, Morgan and colleagues (1994) recruited preservice teachers who were "low-performing" to participate in coaching activities.

**Table 1.** Literature Review of Coaching Studies

Reference	Setting/participants	Evidence-based practice coached	Coaching model	Design and dependent measure	Results/findings
Filcheck et al. (2004)	1 in-service teacher and 1 paraprofessional 17 preschool children (i.e., four- to five-year-olds) with high levels of disruptive behavior	ClassWide Positive Behavior Support	Side-by-side coaching One instructive training session, in-vivo coaching, modeling, and immediate feedback	Single-Subject ABACC design Frequency of inappropriate behaviors per child per minute Frequency of labeled praises exhibited by the teacher per minute Frequency of criticisms exhibited by the teacher per minute	Amount of inappropriate behavior exhibited in the classroom decreased after coaching Teacher used more labeled and unlabeled praises, and fewer criticisms after coaching
Jager et al. (2002)	12 in-service general educators (control and intervention) 10- to 11-year-old students	Direct Instruction in Reading Comprehension	Supervisory 5 in-services plus 3 individual coaching sessions (observation plus feedback)	Quasi-experimental pretest/posttest control group design Gain scores on a High Inference observation (i.e., 5-point rating instrument measuring teacher modeling, guided practice, clear presentation of content)	Teachers trained to use direct instruction showed significantly more characteristics of direct instruction than the control group Some components of direct instruction were more difficult for teachers to implement, even with extensive training (i.e., whole-class engagement, individual feedback, summarizing lesson, and evaluation of student learning)



Kohler et al. (1997)	4 in-service elementary teachers and their classes 1 coach-teacher with 32 years experience and training in coaching process	Direct Instruction plus ClassWide Peer Tutoring	Supervisory All-day in-service plus seven 30- to 45-minute coaching sessions per teacher, during which coach provided nonevaluative feedback	Multiple-Baseline-Across-Teachers Descriptive design for students Teacher Measure: Percentage of correctly implemented Primary Instructional Components (11-item checklist used to rate teachers' conduct of the minilesson, facilitation of cooperative learning, procedures used for closure, and rewards given for on-task behavior) Student Measure: Active engagement	Functional relationship between coaching and higher percentages of Primary Instructional Components completed correctly Percentage of 10-second intervals with active student engagement decreased during baseline to coaching for minilessons, increased slightly from baseline to coaching for peer tutoring, and increased from baseline to coaching for lesson closure
Kohler et al. (1999)	3 in-service kindergarten teachers and their classes	ClassWide Peer Tutoring in Language Arts	Peer Coaching Teachers observed each other implementing minilessons 10- to 20-minute meeting to discuss and evaluate lesson using 2 questionnaires (i.e., student engagement and teacher instruction)	Multiple-Baseline-Across-Teachers Teacher Measures: Statements to focal students during partner activities Refinements in the task that focal students completed Focus and quality of collaboration with coach during 2 coaching phases Student Measures: Social interaction and active engagement	Functional relationship between coaching and: Increased use of suggestions, prompts, questions, and related talk to facilitate students' social interaction with peers Increased adaptation to academic materials Students' social interaction/engagement increased

(continued)

**Table 1. (continued)**

Reference	Setting/participants	Evidence-based practice coached	Coaching model	Design and dependent measure	Results/findings
Kretlow et al. (2009)	3 in-service kindergarten teachers with previous experience teaching Direct Instruction programs at Title I school and their classes, with large population of students at risk for failure	Strategies to increase Active Student Responding (i.e., choral responding, response cards, model-lead-test scaffolding, systematic error correction)	Combination of supervisory and side-by-side I half-day in-service with live and video demonstration and opportunities to practice in pairs with feedback from trainer I preconference, in vivo coaching session, and I postconference per teacher	Multiple-Baseline-Across-Teachers Percentage of correctly implemented GIU per 10 minutes of whole-class math lesson	Teachers' percentage of correctly implemented GIU increased from baseline to post-in-service phase, then increased again during postcoaching phase
Kretlow et al. (in press)	4 in-service first-grade teachers with previous experience teaching Direct Instruction programs at a suburban school with high numbers of English language learners	Strategies to increase Active Student Responding (i.e., choral responding, response cards, model-lead-test scaffolding, systematic error correction)	Combination of supervisory and side-by-side I half-day in-service with live and video demonstration and opportunities to practice in pairs with feedback from trainer I preconference, in vivo coaching session, and I postconference per teacher	Multiple-Baseline-Across-Teachers Percentage of correctly implemented GIU per 10-minute, whole-class numeracy lesson Percentage of correctly implemented GIU per 10-minute, whole-class, uncoached problem solving lesson (i.e., generalization measure) Number of group responses per session	Teachers' percentage of correctly implemented GIU increased from baseline to post-in-service phase, then increased again during postcoaching phase Teachers' generalized improvements in instruction from numeracy to problem solving during coaching phase



Lignuaris-Kraft & Marchand-Martella (1993)	8 preservice special educators, elementary education practicum	Direct Instruction (i.e., Reading Mastery)	Supervisory, more advanced graduate students coached peers	Single-Group Pretest/Posttest with repeated measures	Repeated measures ANOVA showed significant improvements for presentation, signal, and error correction
Maheady et al. (2004)	10 randomly selected preservice general educators	ClassWide Peer Tutoring (i.e., spelling)	Side-by-side (i.e., 2-hour in-service plus follow-up coaching)	No formal student data	All preservice teachers reached high levels of accuracy ( $M = 87.7$ )
Miller et al. (1991)	207 second-through fourth-graders Elementary education practicum	Learning Strategies/Strategic Instruction Model	Supervisory Intermittent 15-minute observations, plus 2 coaching sessions	Single-group pretest/posttest (teachers and students)	Students' weekly spelling test performance increased from precoaching to postcoaching by approximately 25%
Morgan et al. (1994)	6 in-service graduate special education teachers enrolled in summer training on learning strategies	Direct Instruction (i.e., Reading Mastery and Spelling Mastery)	Supervisory 30- to 45-minute meetings twice per week, coaches and students watched videotaped teaching sessions	Multiple-Baseline-Across-Teachers	Functional relationship between coaching sessions and increase in effective teaching behaviors/ decrease in ineffective teaching behaviors
	5 low-performing preservice special educators			Number of effective and ineffective teaching behaviors, measured by the FPMS	Low level of percentage of correct instructional trials in baseline across all four teachers (i.e., 0%–60%)
	Direct Instruction course/practicum			Percentage of correct instructional trials in reading sessions	Functional relationship between coaching and improved instructional
	Coaches, 3 female undergraduate students with			Percentage of correct instructional trials behaviors in spelling	

(continued)

**Table 1. (continued)**

Reference	Setting/participants	Evidence-based practice coached	Coaching model	Design and dependent measure	Results/findings
	superior performance in a Direct Instruction practicum 21 first- through fourth-graders with high-incidence disabilities		Coach and students separately scored DVs, compared, coach modeled as needed, identified target behaviors for improvement	Rate of praise statements and correct responses Number of lessons mastered by groups of students	trials (i.e., 60%–100%), stable change in level and trend across all teachers Rates of praise and student responses increased
Peck et al. (1989)	3 preschool teachers with limited experience working with students with disabilities Inclusive preschool with 60–70 students enrolled	Systematic Prompting related to IEP objectives during ongoing instruction (i.e., modeling, gesturing, physical guidance)	Supervisory Coach video-recorded teaching sessions, then reviewed clips with teachers using guiding questions about prompting strategies	Multiple-Baseline-Across-Teachers and Multiple-Baseline-Across-Students Teacher measure: Rate per minute of teacher instructional behaviors (i.e., IEP-related instructional prompts of teachers) Student measure: Rate of IEP-targeted responses by students	Increases in targeted behavior (i.e., prompts) for all 3 teachers during coaching Increases in targeted behavior for 2 teachers during generalization Changes in IEP-targeted responses highly correlated with changes in teacher behavior; rate of IEP responses improved for all students Supervisory and peer coaching both resulted in a significant increase in effective behaviors and decrease in ineffective behaviors
Pierce & Miller (1994)	32 preservice special educators Practicum course in mild intellectual disabilities/self-contained	Direct Instruction (Rosenshine, 1983; across varied content areas)	Supervisory (control) and peer coaching (intervention) 50-minute weekly seminar plus	2 × 2 mixed design with 1 between factor and 1 within factor Number of effective and ineffective teaching behaviors, measured by the FPMS,	

<p>contained classrooms at the elementary and secondary levels</p> <p>University supervisor with master's degree and classmates served as coaches</p> <p>8 elementary general educators at a Title I school with established School-Wide Positive Behavior Support model</p> <p>Stitcher et al. (2006)</p>	<p>conferences and coaching support group</p> <p>including lesson presentation, instructional organization, and verbal communication, and classroom management</p> <p>No formal student data were collected</p>	<p>Increased opportunities to respond to academic instruction</p> <p>Supervisory 2 in-services plus 1 preconference, 1 observation, and 1 postconference</p> <p>Multiple-Baseline-Across-Teachers</p> <p>Teacher measures: Percentage of correctly implemented opportunities to respond compared to optimal levels from the literature</p> <p>Student Measures: 16 target students</p> <p>Mean difference from baseline to post-in-service and coaching on: Permanent literacy product, standard literacy scores on districtwide literacy assessments</p> <p>Number of office referrals</p>	<p>In-service plus coaching led to improvements in some aspects of increased OTR, but not all: 5 out of 8 teachers met criterion goal for instructional talk, 4 of 8 for feedback, 2 of 8 for wait time, none for prompting</p> <p>Student outcomes varied widely</p>
--	---	---	---

Note: FPMS = Florida Performance Measurement System; GIU = group instructional units; IEP = Individualized Education Program; OTR = Opportunity to Respond.

## Coaches

Individuals who provided coaching to teachers had a wide range of characteristics. In several studies, researchers (i.e., university faculty, doctoral student, university student teacher supervisor) provided direct training to teachers including in-service or course work along with follow-up individualized coaching (Filcheck et al., 2004; Kohler et al., 1997; Kretlow et al., 2009, in press; Maheady et al., 2004; Miller et al., 1991). In other studies, an experienced teacher provided coaching (Hasbrouck, 1997; Kohler et al., 1999; Pierce & Miller, 1994). In a few studies, the researcher provided the initial in-service or preservice training, but a second individual (e.g., experienced teacher, fellow student) provided the coaching (Lignugaris-Kraft & Marchand-Martella, 1993; Morgan et al., 1994; Stitche et al., 2006). Peck and colleagues (1989) did not describe the coach other than as a "facilitator."

*Coaching characteristics.* Although all teachers received a form of individualized follow-up support, the nature of specific coaching activities varied widely. The majority of studies included a combination of in-service (or course work) followed by individualized coaching sessions (Filcheck et al., 2004; Hasbrouck, 1997; Kohler et al., 1997, 1999; Kretlow et al., 2009, in press; Maheady et al., 2004; Miller et al., 1991; Morgan et al., 1994). In the remaining studies, coaching activities began with an observation followed by coaching sessions (Kohler et al., 1999; Lignugaris-Kraft & Marchand-Martella, 1993; Peck et al., 1989; Pierce & Miller, 1994). The total duration of professional development activities ranged from several hours to 16 weeks.

We found two primary methods of coaching consistent with the conceptual literature in the studies reviewed: supervisory coaching and side-by-side (i.e., in vivo) coaching (Joyce & Showers, 1995; Raney & Robbins, 1989). Across studies, supervisory coaching elements included: (a) in-service or multiple observations of the teacher by the coach prior to intervention; (b) a consultative "feedback"

meeting; (c) follow-up observations; and (d) at least one more consultative "feedback" meeting (Hasbrouck, 1997; Kohler et al., 1997; Lignugaris-Kraft & Marchand-Martella, 1993; Miller et al., 1991; Morgan et al., 1994; Peck et al., 1989; Pierce & Miller, 1994; Stitche et al., 2006).

In the side-by-side studies, coaching elements included (a) in-service or multiple observations of the teacher by the coach prior to intervention; (b) consultative "feedback" meeting; (c) demo lesson taught by the coach in teachers' classrooms; (d) follow-up observations; and (e) at least one more consultative "feedback" meeting (Filcheck et al., 2004; Kohler et al., 1997, 1999; Kretlow et al., 2009, in press; Maheady et al., 2004; Peck et al., 1989). In addition, Stitche and colleagues (2006) provided ongoing email assistance to teachers from the faculty member providing coaching. None of the other studies directly stated whether or not teachers had access to coaches in between coaching activities to request assistance.

*Coaching fidelity.* Only four studies collected data on the extent to which coaches adhered to a specific protocol for coaching activities (Hasbrouck, 1997; Kretlow et al., 2009, in press; Morgan et al., 1994). The percentage of sessions during which fidelity data were collected ranged from 13% to 100%. Across all four studies, a second observer watched either live or video-recorded in-service and coaching sessions and marked the occurrence of specific activities on a checklist. Coaching fidelity was calculated using the point-by-point method.

## Dependent Variables

All 13 studies used a measure of teaching accuracy as the primary dependent variable; however, accuracy was defined differently across studies. The dependent variables can be separated into two categories: those that were specific to the evidence-based practice coached (i.e., essentially a measure of fidelity); and those that were not specific to the evidence-based practice, but included specific

components authors defined as “effective instruction” (e.g., presentation of content, materials, classroom management).

For example, Lignugaris-Kraft and Marchand-Martella (1993) coached 19 undergraduate special education students to deliver Direct Instruction (DI) programs (e.g., Reading Mastery, Corrective Reading, Spelling Mastery) to students with intellectual disabilities. The authors’ dependent variable was a frequency count of specific DI components teachers delivered accurately in lessons including: signals (e.g., holding one finger up for each sound during phonemic awareness activities); presentation accuracy (e.g., pointing to letters, using model-lead-test procedures); specific praise; and fast pacing. Five other studies in this review used a dependent variable that was specifically tied to the practice coached (Filcheck et al., 2004; Kohler et al., 1997; Maheady et al., 2004; Morgan et al., 1994).

In contrast, Miller and colleagues (1991) provided an example of a more “general” dependent variable. Researchers in this study coached six in-service special education teachers to use three of the Learning Strategies curricula: Paraphrasing, Sentence Writing, and Test Taking. The authors’ dependent variable did *not* include specific components of these programs, but instead included frequency counts of effective and ineffective overall teaching behaviors that occurred while teachers were implementing the programs (e.g., organization, presentation strategies, communication with students, classroom management). Seven other studies in this review used similar “general” characteristics of teaching as dependent variables (Hasbrouck, 1997; Kohler et al., 1999; Kretlow et al., 2009, in press; Peck et al., 1989; Pierce & Miller, 1994).

All 13 studies showed that coaching interventions led to improvements in teaching accuracy. In all nine of the single-subject studies, effects were demonstrated in a multiple-baseline-across-participants design. Across these studies, coaching resulted in improvements in level or trend, or decreased variability, and in some cases, level, trend, and

variability improved. In the single-subject studies that collected data during in-service trainings occurring prior to coaching, teachers showed initial improvement in accuracy post-in-service with a second, higher level of accuracy after coaching (Kretlow et al., 2009, in press). In the four studies using group designs, improvements in teaching accuracy were significant at the .05 level for one study and at the .001 level for the other three. Only one study reported effect size (Hasbrouck, 1997), which was .41, considered a moderate to high effect size in the educational literature (Tabachnick & Fidell, 2007).

### *Teacher Perceptions*

Eight studies collected social validity data examining the extent to which teachers found coaching activities beneficial (Filcheck et al., 2004; Hasbrouck, 1997; Kohler et al., 1997, 1999; Kretlow et al., 2009, in press; Morgan et al., 1994; Peck et al., 1989). The majority of studies used a Likert-type scale or questionnaire to evaluate social validity. Kohler and others (1999) and Peck and colleagues (1989) conducted in-depth interviews with teachers after the coaching process concluded. Teachers across all eight studies positively rated coaching activities. The most common suggestion for improvement of coaching activities was more individual coaching sessions.

### *Student Outcomes*

Eight studies reported a student outcome (Filcheck et al., 2004; Kohler et al., 1997, 1999; Kretlow et al., in press; Maheady et al., 2004; Morgan et al., 1994; Peck et al., 1989; Stitche et al., 2006). The most frequently reported student outcome variable was academic engagement or on-task behavior. Two studies examined an academic outcome; Maheady and others (2004) reported weekly spelling test performance, and Stitche and colleagues (2006) reported student performance on teacher- and district-created literacy measures.

## Discussion

The purpose of this review was to summarize studies that examined the effects of coaching on improvements in preservice and in-service teachers' implementation of evidence-based practices. The collection of studies we reviewed provides strong evidence for the effectiveness of coaching in promoting the fidelity of evidence-based practices.

### *Implications for Preservice and In-Service Teacher Preparation*

Given the well-documented research-to-practice gap in education (Greenwood & Abbott, 2001), the literature identified in this review offers some critical implications for preservice and in-service teacher preparation. Specifically, the studies reviewed suggest common themes for (a) critical components of coaching, (b) measurement of teacher behaviors, and (c) measurement of correlated student achievement.

*Critical components of coaching.* First, studies identified in this review used supervisory coaching, side-by-side coaching, or a combination of both. Within those broad categories, some critical components of coaching were present in almost all studies including (a) highly engaged, instructive group training sessions; (b) follow-up observation(s); and (c) specific feedback, often including sharing observation data and self-evaluation followed by modeling (i.e., in vivo or outside the classroom). Almost all interventions included a combination of some form of instructive training and individualized follow-up coaching. Specifically, teachers were often grouped together for initial course work or in-service trainings, during which an instructor (i.e., someone who sometimes served as the coach, but not always) provided an overview of the particular strategy and opportunities for teachers to practice in simulated activities (e.g., take turns teaching one another). With the exception of one study (Pierce & Miller, 1994), these sessions involved relatively small groups of teachers (i.e., 12 or fewer), especially when

compared to typical undergraduate/graduate courses and in-service professional development opportunities. Working with a small number of teachers on targeted skills likely allowed trainers to give more opportunities to practice and more individualized feedback.

Second, all 13 interventions included multiple observations with subsequent feedback, ranging from two live observations to daily observations for several months (i.e., audio-recorded). Observation is likely a key component of promoting the use of evidence-based practices to the classroom because it provides a setting event for the particular teacher behaviors to occur (Cooper, Heron, & Heward, 2007). To put it simply, the fact that teachers knew they would be regularly observed, "set up" the occasion for them to implement the practice more regularly. Regular observation is rarely a typical activity in preservice training (i.e., with the exception of student teaching; Scheeler et al., 2009) and even more rarely a typical activity after in-service training (Desimone et al., 2002; Yoon et al., 2007). Using audio/video technology to record teacher practices is one way to increase repeated observation. Using technology could allow teacher educators to assess teacher performance during and after training activities and also allow an assessment of sustainability.

Third, coaches across all 13 studies offered individualized feedback derived directly from data taken during observations. Feedback took many forms, including the teacher and coach evaluating video recordings of lessons together, direct stating of instructional strengths and opportunities for improvement, reciprocal peer observation and feedback, and in vivo feedback (i.e., coach directly intervenes in a nonevaluative manner during lesson). In addition, whether feedback occurred during or after instruction, coaches frequently provided modeling of specific instructional skills, followed by immediate opportunities for teachers to practice the skill again. Previous research has shown that this "my turn-your turn" approach to error correction improves acquisition of new skills for students (Barbetta, Heron, & Heward, 1993; Barbetta, Heward,

Bradley, & Miller, 1994; Gettinger, 1993; Nelson, Alber, & Gordy, 2004); it appears the same may be true for teachers.

*Measuring changes in teacher performance.* Overwhelmingly, coaching improved the accuracy of teaching behaviors across studies reviewed. Researchers in all studies measured teaching improvements by capturing a combined set of discrete skills, then calculating the overall number or percentage of accuracy—essentially, a measure of instructional fidelity. Some studies targeted instructional skills specific to the evidence-based strategy teachers were using (e.g., signaling in Direct Instruction, awarding points during ClassWide Peer Tutoring, applying reinforcement in Positive Behavior Support), whereas others attempted to capture critical instructional design variables that could be applied across teaching techniques (e.g., modeling, prompting, engagement). Two measures with some previous validation were used across a few studies: the Florida Performance Measurement System (Miller et al., 1991; Pierce & Miller, 1994) and the instructional trial (Kretlow et al., 2009, in press; Morgan et al., 1994).

Regardless of the specific measure used, the majority of studies identified a few components of instruction as salient: (a) presentation of new skills, including modeling and systematic prompting; (b) guided practice, including multiple opportunities to respond; and (c) active engagement. These salient instructional design components suggest the utility of a universal measure of instructional proficiency. Such a measure could be a helpful tool in making coaching a more transportable training practice. For example, if a universal measure could be validated to capture critical components of instruction that crossed evidence-based practices, evaluation materials could be more widely disseminated and used in schools.

*Measuring changes in student performance.* In a previous meta-analysis of professional development studies, Yoon et al. (2007) reported that for any professional development to be deemed effective, it must demonstrate

observable changes in teaching practice and student achievement. Only five studies in this review reported any student data, and of those five only three demonstrated experimental effects of coaching and improvements in student outcomes (Kohler et al., 1997, 1999; Maheady et al., 2004; Morgan et al., 1994; Peck et al., 1989). Student outcomes in these studies were measured by (a) weekly spelling test performance, (b) rate of correct student responses during lessons, (c) frequency of inappropriate behaviors, (d) mean 10-second intervals of student engagement, (e) rate of Individualized Education Program targeted responses, and (f) teacher-created literacy permanent products and district-created literacy assessments.

None of the studies prompted teachers to use student data to make instructional decisions. Other studies have demonstrated the importance of training teachers to use data-based decision making to help sustain the use of evidence-based practices (Capizzi & Fuchs, 2005). It is logical that teachers may be more likely to continue implementing a practice with fidelity if they can “see” real changes in student achievement as a result of their instruction. Similarly, teachers may be more willing to modify instruction if they can “see” that students are not progressing academically.

### *Implications for Practice*

The results of this review suggest coaching is a promising practice for promoting high fidelity of evidence-based practices from training settings to real classroom settings. Data across studies suggest that group instruction, such as in-service and college class sessions, are typically not sufficient to produce instructional accuracy with a newly learned practice in the classroom. Therefore, preservice and in-service teacher training should include a coaching component whenever possible, to intentionally train teachers to use evidence-based practices in the classroom. We propose that adding a coaching component to teacher training would not be particularly difficult, given that the coaching



interventions used across studies were not time intensive or costly. In addition, several studies showed substantial improvements in instructional proficiency after only one or two coaching sessions, which suggests coaching is a manageable technique for the instructor.

*General implications for training activities.*

The studies reviewed offer some general guidelines for developing course activities and professional development sessions, as well as specific coaching components. First, when teachers are learning a new instructional practice, it is advantageous to use a small-group format for instructive sessions (i.e., class sessions, in-services at schools). Within the small groups, teacher educators should provide multiple opportunities to practice the instructional strategy with specific praise and with error correction including modeling, when necessary. It may be helpful to tell teachers that an observation will follow the initial training, to prompt them to try out the strategy in the classroom (i.e., they will use it because they know someone is going to observe at some point). Second, prior to coaching, teacher educators should conduct at least one observation (e.g., live, video-, or audio-recorded) to determine with which specific skills the teacher is having difficulty. Then, feedback and coaching activities should directly target skills that need to be firmed up. Coaching sessions should include some form of modeling, whether it is during an actual lesson or one-on-one with the teacher. Teachers may also benefit from self-evaluation of lesson effectiveness, which should then be compared with the coach's evaluation.

*Implications for preservice teacher education.*

This review suggests that preservice teachers need (a) high-quality instructive training, including opportunities to "simulate" instruction during course work with feedback from the instructor; (b) multiple opportunities to practice newly learned strategies with real students; and (c) individualized observation, feedback, and modeling, including side-by-side or supervisory coaching whenever possible. Simulated teaching opportunities during in-class sessions may be particularly important

for preservice teachers because most often they do not have classroom experiences with which to ground their new knowledge. In practicum and student teaching experiences, both an experienced classroom teacher and a faculty supervisor could serve as coaches for a teacher candidate. Coaching may be an easy replacement for the typical "observe and give feedback" format used in field experience, and may make these experiences even more powerful, given that the coaching would occur in the generalization setting (i.e., the classroom).

*Implications for in-service teacher education.*

In-service teachers typically receive instruction during professional development activities, such as workshops, seminars, and conferences. Other research clearly shows that these kinds of activities do not lead to sustainable changes in instruction (Yoon et al., 2007). However, the conundrum remains that professional development is one of the only ways to impart new knowledge to practicing teachers. The good news is that results of this review show that redesigning professional development activities could make them more likely to result in observable changes. First, whenever possible, in-services should be delivered in the teacher's context (e.g., in the school, with colleagues, applying new learning to existing materials). Second, trainers should deliver instruction to small groups of teachers whenever possible, to provide multiple opportunities to practice new instructional techniques with corrective feedback from the trainer. For example, when introducing a Tier 2 intervention for Responsiveness to Intervention in reading, a literacy facilitator might provide in-service training to the tutors delivering the instruction and the general educators for only one grade level at a time (e.g., 8–10 people). Given the small-group format, she or he could put the tutors and teachers in small groups, simulate an instructional setting, watch the participants try the strategy, and provide targeted feedback and/or modeling when needed. In contrast, with a group of 50 teachers representing all grade levels, it would be much more

difficult for a literacy facilitator to monitor all participants and give enough opportunities to practice. Third, the in-service should be followed by observations. We recognize that many times the trainer is not a staff member and therefore may not be able to conduct observations. However, in most schools, some form of instructional support exists (e.g., literacy facilitator, math coach, lead teacher). These personnel could provide follow-up observations and coaching after an initial in-service provided by a nonstaff member.

### *Implications for Future Research*

The 13 studies summarized in this review clearly demonstrated that coaching can improve the fidelity of a variety of evidence-based practices, from initial training to classroom settings. Although evidence for coaching for Direct Instruction, Learning Strategies, Positive Behavior Support, and Active Student Responding exists, future research needs to examine the impact of coaching across other well-researched practices and programs. In addition, the studies we reviewed focused primarily on elementary teachers. Only two studies were conducted in middle or secondary school settings.

In addition, only three studies experimentally demonstrated positive effects of coaching on improvements in student performance. Before wholeheartedly advocating the use of coaching to improve instruction, future research must examine the impact of coaching on changes in student performance using reliable and valid measures. For example, studies should examine the impact of specific coaching models on changes in students' reading performance using psychometrically valid curriculum-based measures (CBM) such as the Dynamic Indicator of Basic Early Literacy Skills (Good & Kaminski, 2002) or math and writing performance using CBM such as AIMSweb. Such studies would determine whether coaching is salient enough to produce lasting changes in student achievement on measures that are predictive of later

school success. Those conducting these studies need to use research designs that will lead to external validity. In addition, if research can identify the level of fidelity (e.g., 90%, 95%) at which students benefit substantially, coaching could become criterion referenced. Knowing the levels of fidelity that predict student achievement could potentially make the process of determining who needs the most support more efficient. Correlation or regression analysis could be used to identify specific levels of fidelity that lead to critical student gains and to determine if student performance continues to improve as instructional accuracy improves (i.e., determine if a ceiling in student effects exists). Current research has not answered the question of how professional development impacts student achievement (Wayne, Yoon, Zhu, Cronen, & Garet, 2008).

An equally important variable to investigate is the effect of coaching on the sustainability of changes in teaching practices. Some studies in this review showed that changes were sustained for up to three months (Kretlow et al., 2009, in press), but during this time, teachers were "observed" daily via audio recording, which likely prompted them to use the strategies more regularly. Studies should investigate whether coaching produces changes that last after the follow-up support and frequent observation and/or data collection is removed (e.g., 6 months, 1 year).

It is also critical to know who benefits most from coaching and how much coaching is needed to produce lasting changes in teaching performance and salient improvements in student performance. The studies in this review provided a wide range of coaching sessions (i.e., from one session to weekly sessions for one year), as well as a wide range of teacher characteristics (i.e., low-performing teachers to seasoned, high-performing teachers). Comparative studies could help identify the most parsimonious coaching model—that is, the fewest in-service and coaching sessions that will lead to gains in teacher practice and student achievement. Such studies could also identify the particular coaching model

that is most effective (i.e., supervisory vs. side-by-side).

Teacher preferences should also be considered in the overall effectiveness of coaching models. Survey or qualitative studies could help identify the variables of supervisory or side-by-side coaching that teachers find more acceptable and helpful. In-depth interviews with teachers, such as the ones conducted by Kohler and colleagues (1999) and Peck and others (1989) may produce more useful information than a social validity questionnaire, as the researcher or coach can ask open-ended and clarifying questions. Teacher satisfaction with professional development models is critical to their long-term success (Wayne et al., 2008).

In summary, the body of evidence supporting coaching is emerging. The 13 studies reviewed here are consistent with the professional development literature in that they demonstrated the importance of follow-up support. Hall and Hord (2006) indicated teacher change is not an event but rather a process by which individuals gradually move to become skilled in the implementation of new strategies. The use of coaching in teacher preparation takes this into consideration as it moves beyond a "one-shot" training to guide teachers toward fidelity in the classroom over time. The persisting research-to-practice gap, along with the importance of improving fidelity, are salient reasons to revisit current pre-service and in-service training models. This review offers one effective innovation for teacher educators.

### Declaration of Conflicting Interests

The authors declared no conflicts of interest with respect to the authorship and/or publication of this article.

### Funding

The authors received no financial support for the research and/or authorship of this article.

### References

References marked with an asterisk indicate studies included in the meta-analysis.

- Abbott, M., Walton, C., Tapia, Y., & Greenwood, C. (1999). Research to practice: A blueprint for closing the gap in local schools. *Exceptional Children, 65*, 339-352.
- Allen, D. W., & LeBlanc, A. C. (2004). *Collaborative peer coaching that improves instruction: The 2 + 2 performance appraisal model*. Thousand Oaks, CA: Corwin.
- Barbetta, P. M., Heron, T. E., & Heward, W. L. (1993). Relative effects of whole-word and phonetic-prompt error correction on the acquisition and maintenance of sight words by students with developmental disabilities. *Journal of Applied Behavior Analysis, 26*, 99-110.
- Barbetta, P. M., Heward, W. L., Bradley, D. M., & Miller, A. D. (1994). Effects of immediate and delayed error correction on the acquisition and maintenance of sight words by students with developmental disabilities. *Journal of Applied Behavior Analysis, 27*, 177-178.
- Bartholomew, C. C., & Kretlow, A. G. (2010). *Teacher reported barriers to providing high quality reading instruction in kindergarten, first, and second grades*. Manuscript in preparation.
- Blakely, M. R. (2001). A survey of levels of supervisory support and maintenance of effects reported by educators involved in direct instruction implementation. *Journal of Direct Instruction, 1*, 73-83.
- Bulgren, J. A., Lenz, B. K., McKnight, M., Davis, B., Grossen, B., Marquis, J., . . . & Shumaker, J. B. (2002). *The educational context and outcomes for high school students with disabilities: The perceptions of general education teachers. Report to the Office of Special Education Programs, U.S. Department of Education by the Institute for Academic Access* (Research Report No. 7). Retrieved from ERIC database. (ED469287)
- Burns, M. K., & Ysseldyke, J. E. (2009). Reported prevalence of evidence-based instructional practices in special education. *The Journal of Special Education, 43*, 3-11.
- Buzhardt, J., Greenwood, C. R., Abbott, M., & Tapia, Y. (2007). Scaling up ClassWide Peer Tutoring: Investigating barriers to wide-scale implementation from a distance. *Learning Disabilities—A Contemporary Journal, 5*(2), 75-96.

- Caffarella, R. S., & Barnett, B. G. (1994). Characteristics of adult learners and foundations of experimental learning. *New Directions for Adult and Continuing Education*, 62, 29-42.
- Capizzi, A. M., & Fuchs, L. S. (2005). Effects of curriculum-based measurement with and without diagnostic feedback on teacher planning. *Remedial and Special Education*, 26, 159-174.
- Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis*, (2nd ed.). Upper Saddle River, NJ: Pearson/Merrill-Prentice Hall.
- Darling-Hammond, L., Wei, R. C., Andree, A., Richardson, N., & Orphanos, S. (2009). State of the profession: Study measures status of professional development. *Journal of Staff Development*, 30(2) 42-50.
- Desimone, L. M., Porter, A. C., Garet, M. S., Yoon, K. S., & Birman, B. F. (2002). Effects of professional development on teachers' instruction: Results from a three-year longitudinal study. *Education Evaluation and Policy Analysis*, 24, 81-112.
- DiGennaro, F. D., Martens, B. K., & Kleinmann, A. E. (2007). A comparison of performance feedback procedures on teachers' treatment implementation integrity and students' inappropriate behavior in special education classrooms. *Journal of Applied Behavior Analysis*, 40, 447-461.
- Elmore, R. F. (2006). *School reform from the inside out: Policy, practice, and performance*. Cambridge, MA: Harvard Education Press.
- \*Filcheck, H. A., McNeil, C. B., Greco, L. A., & Bernard, R. S. (2004). Using a whole-class token economy and coaching of teacher skills in a pre-school classroom to manage disruptive behavior. *Psychology in the Schools*, 41, 351-361.
- Forness, S. R. (2001). Special education and related services: What have we learned from meta-analysis? *Exceptionality*, 9, 185-197.
- Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Ferguson, C. (1992). Effects of expert system consultation within curriculum-based measurement, using a reading maze task. *Exceptional Children*, 58, 436-450.
- Furtak, E. M., Ruiz-Primo, M. A., Shemwell, J. T., Ayala, C. C., Brandon, P. R., Shavelson, R. J., & Yue, Y. (2008). On the fidelity of implementing embedded formative assessments and its relation to student learning. *Applied Measurement in Education*, 21, 360-389.
- Gettinger, M. (1993). Effects of error correction on third graders' spelling. *Journal of Educational Research*, 87, 39-45.
- Good, R. H., & Kaminski, R. A. (Eds.). (2002). *Dynamic indicators of basic early literacy skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement.
- Gordon, S. P. (2004). *Professional development for school improvement*. Boston, MA: Pearson Education.
- Greenwood, C. R., & Abbot, M. (2001). The research to practice gap in special education. *Teacher Education and Special Education*, 24, 276-289.
- Hall, G. E., & Hord, S. M. (2006). *Implementing change: Patterns, principles, and potholes* (2nd ed.). Boston, MA: Pearson Education.
- Hasbrouck, J. E. (1997). Mediated peer coaching for training preservice teachers. *Journal of Special Education Teachers*, 31, 251-271.
- Institute of Education Sciences. (2007). What Works Clearinghouse intervention report: ClassWide Peer Tutoring. Retrieved from [http://ies.ed.gov/ncee/wwc/pdf/WWC\\_CWPT\\_070907.pdf](http://ies.ed.gov/ncee/wwc/pdf/WWC_CWPT_070907.pdf)
- \*Jager, B., Reezigt, G. J., & Creemers, B. P. (2002). The effects of teacher training on new instructional behavior in reading comprehension. *Teaching and Teacher Education*, 18, 831-842.
- Joyce, B., & Showers, B. (1995). *Student achievement through staff development: Fundamentals of school renewal*. New York, NY: Longman.
- Kavale, K. A., & Forness, S. R. (1999). Effectiveness of special education. In C. R. Reynolds & T. B. Gutkin (Eds.), *The handbook of school psychology* (3rd ed., pp. 984-1024). New York, NY: Wiley.
- Klingner, J. K., Vaughn, S., Hughes, M. T., & Arguelles, M. E. (1999). Sustaining research-based practices in reading: A 3-year follow-up. *Remedial and Special Education*, 20, 263-274.
- \*Kohler, F. W., Crilley, K. M., Shearer, D. D., & Good, G. (1997). Effects of peer coaching on teacher and student outcomes. *The Journal of Educational Research*, 90, 240-250.
- \*Kohler, F. W., Ezell, H. K., & Paluselli, M. (1999). Promoting changes in teachers' conduct

- of student pair activities: An examination of reciprocal peer coaching. *The Journal of Special Education*, 33, 154-165.
- Kovaleski, J. F., Gickling, E. E., Morrow, H., & Swank, P. R. (1999). High versus low implementation of instructional support teams: A case for maintaining program fidelity. *Remedial and Special Education*, 20, 170-183.
- \*Kretlow, A. G., Cooke, N. L., & Wood, C. L. (in press). Using inservice and coaching to increase the accurate use of research-based strategies. *Remedial and Special Education*.
- \*Kretlow, A. G., Wood, C. L., & Cooke, N. L. (2009). Using in-service and coaching to increase kindergarten teachers' accurate delivery of group instructional units. *Journal of Special Education*. Advance online publication. doi: 10.1177/0022466909341333
- \*Lignugaris-Kraft, B., & Marchand-Martella, N. (1993). Evaluation of preservice teachers' interactive teaching skills in a Direct Instruction practicum using student teachers as supervisors. *Teacher Education and Special Education*, 16, 309-318.
- \*Maheady, L., Harper, G. F., Mallette, B., & Karnes, M. (2004). Preparing preservice teachers to implement class wide peer tutoring. *Teacher Education and Special Education*, 27, 408-418.
- McLeskey, J., & Billingsley, B. S. (2008). How does the quality and stability of the teaching force influence the research-to-practice gap? A perspective on the teacher shortage in special education. *Remedial and Special Education*, 29, 293-305.
- Merriam, S. B. (2001) Andragogy and self-directed learning: Pillars of adult learning theory. *New Directions for Adult and Continuing Education*, 89, 3-13.
- \*Miller, S. P., Harris, C., & Watanabe, A. (1991). Professional coaching: A method for increasing effective and decreasing ineffective teacher behaviors. *Teacher Education and Special Education*, 14, 183-191.
- \*Morgan, R. L., Menlove, R., Salzberg, C. L., & Hudson, P. (1994). Effects of peer coaching on the acquisition of Direct Instruction skills by low-performing teachers. *The Journal of Special Education*, 28, 59-76.
- Nelson, J. S., Alber, S. R., & Gordy, A. (2004). Effects of systematic error correction and repeated readings on the reading accuracy and proficiency of second graders with disabilities. *Education and Treatment of Children*, 27, 186-198.
- Office of Special Education Programs. (2010). *Is school-wide positive behavior support an evidence-based practice?* Retrieved from <http://pbis.org/research/default.aspx>
- O'Reilly, M. G., & Renzaglia, A. (1992). Teaching systematic instruction competencies to special education student teachers: An applied behavioral supervision model. *Journal of the Association for the Severely Handicapped*, 17, 104-111.
- \*Peck, C. A., Killen, C. C., & Baumgart, D. (1989). Increasing implementation of special education instruction in mainstream preschools: Direct and generalized effects of nondirective consultation. *Journal of Applied Behavior Analysis*, 22, 197-210.
- \*Pierce, T., & Miller, S. P. (1994). Using peer coaching in preservice practica. *Teacher Education and Special Education*, 17, 215-223.
- Promising Practices Network. (2005). A meta-analytic review of Direct Instruction programs. Retrieved from <http://www.promisingpractices.net/program.asp?programid=146>
- Randolph, J. J. (2007). Meta-analysis of the research on response cards: Effects on test achievement, quiz achievement, participation, and off-task behavior. *Journal of Positive Behavior Interventions*, 9, 113-128.
- Raney, P., & Robbins, P. (1989). Professional growth and support through peer coaching. *Educational Leadership*, 35, 35-38.
- Rosenshine, B. (1983). Teaching functions in instructional programs. *Elementary School Journal*, 83, 335-351.
- Scheeler, M. C., Bruno, K., Grubb, E., & Seavey, T. L. (2009). Generalizing teaching techniques—From university to K–12 classrooms: Teaching preservice teachers to use what they learn. *Journal of Behavioral Education*, 18, 189-210.
- \*Stitcher, J. P., Lewis, T. J., Richter, M., Johnson, N. W., & Bradley, L. (2006). Assessing antecedent variables: The effects of instructional variables on student outcomes



- through in-service and peer coaching professional development models. *Education and Treatment of Children*, 29, 665-692.
- Swanson, H. L., & Hoskyn, M. (1998). Experimental intervention research on students with learning disabilities: A meta-analysis of treatment outcomes. *Review of Educational Research*, 68, 277-321.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics*. New York, NY: Harper Collins.
- Wayne, A. J., Yoon, K. S., Zhu, P., Cronen, S., & Garet, M. S. (2008). Experimenting with teacher professional development: Motives and methods. *Educational Researcher*, 37, 467-479.
- Yoon, K. S., Duncan, T., Lee, S. W., Scarloss, B., & Shapley, K. L. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (Issues & Answers Report, REL 2007-No. 033). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. Retrieved from [http://ies.ed.gov/ncee/edlabs/regions/southwest/pdf/REL\\_2007033.pdf](http://ies.ed.gov/ncee/edlabs/regions/southwest/pdf/REL_2007033.pdf)

### Bios

**Allison Graves Kretlow** is an assistant professor of exceptional education at James Madison University. She is the author of publications related to instructional coaching, early reading and math intervention, and active engagement strategies. Her current research interest is the development of sustainable in-service teacher training programs for schoolwide models of reading.

**Christina C. Bartholomew** is an assistant professor of exceptional education at James Madison University. She is the author of publications related to professional development, assessment practices, self-determination, and teacher perspectives, including coauthorship of a textbook, *Universal Design for Transition: A Roadmap for Planning and Instruction*, that focuses on combining instructional and transition strategies. Her current research interests include using teacher perspectives to design effective professional development opportunities.