Supervisory Teacher Coaching in K-12 Classrooms

Millicent Carmouche, Ph.D.

Jelisa Thompson, M.Ed.

Alabama Agricultural and Mechanical University

Abstract

Coaching has been identified as a critical support for persistent use of newly adopted practices and skills (Joyce & Showers, 1982). A systematic review of the literature was conducted to examine the literature base on supervisory coaching, an approach in which an outside expert or supervisor gives specific, positive, and corrective coaching when needed and is offered to the teacher after the completion of the observed lesson in an effort to move the recipient toward a desired level of performance (Joyce & Showers, 1981; 1982; Maeda, 2001; Simonsen, Myers, & DeLuca, 2010). Sixteen quasi-experimental and single-subject studies were identified and reviewed using quality indicators specific to quasi-experimental (Gersten et al., 2005) and single subject (Horner et al., 2005) research. Only six of the sixteen studies met all quality indicators. Mixed results were found across the studies, with six reporting improved teacher results and four reporting improved student behaviors.

Key Terms: Supervisory Coaching, K-12 classrooms

Supervisory Teacher Coaching in K-12 Classrooms

In recent years, the push for increased rigor and accountability in kindergarten through twelfth grade (k-12) classrooms has been at the forefront of national, state, and local legislators' educational agendas (National Statistics for Education Statistics, 2007). As legislators and instructional reformers lead the charge to change what children learn and how they learn it, classroom teachers are responsible for the implementation of these changes. In addition to daily classroom responsibilities that include instruction, assessment, classroom management, and differentiation for a variety of learners, teachers are also expected to continue their own learning (Darling-Hammond, 2014). While teachers are expected to be lifelong learners (Merriam, Caffarella, Baumgertner, 2012), their development as professionals must be effective and meaningful. In-service teachers rely overwhelmingly on professional development to learn new evidence-based techniques to implement in the classroom (Desimone, Porter, Garet, Yoon & Birman, 2002). With continual professional development, as well as daily classroom responsibilities, teachers may need additional supports to reinforce professional learning (Gregory, Allen, Mikami, Hafen & Pianta, 2014).

What is Professional Development?

Hargreaves (2014) defines *professional development* (PD) as the "experiences that take place within a collaborative culture of shared leadership, that increase educators' knowledge about content and pedagogy and enable them to use that knowledge to improve classroom and school

practices that improve student learning" (p. 44). This definition describes the importance of the professional growth of the teacher as an individual, but also takes into account the need for teachers to learn as a community and the impact PD has on the entire school community. Researchers have suggested that traditional PD generally consists of three activities: (1) within district workshops, (2) courses for college credit, and (3) out-of-district workshops (Desimone et al., 2002). These activities are conventionally comprised of one-time workshops, in which teachers are inactive consumers of knowledge, with little to no follow-up (Loucks-Horsley, 1998, Putnam & Borko, 2000).

Garet and colleagues (2001) surveyed 1027 teachers in 358 school districts. Teachers were asked to compare recent PD opportunities they had attended as grantees of the Eisenhower Professional Development Program over four semesters. Teachers reported that short term "sit and get" PD delivered by outside experts with little connection to the reality of the classroom did not have the same positive effects as long term hands-on PD. They reported that short term PD rarely translated into actual and prolonged implementation in the classroom.

What Makes Professional Development Effective?

School administrators and PD providers must plan and implement effective PD that engages and encourages teachers and improves the instruction students receive (Billingsley, 2005). What constitutes effective PD has been discussed comprehensively in the literature (Birman, Desimone, Porter & Garet, 2000; Borko, 2004; Leko & Brownell, 2009). Some general principles of effective PD have been suggested over the years. First, in a report to US Department of Education, Yoon and colleagues (2007) recommended that effective PD be coherent and aligned with teachers' objectives and needs in the classroom (Klingner, 2004; Penuel, Fishman, Yamaguchi, & Gallegher, 2007; Yoon, Duncan, Lee, Scarloss, & Shapley, 2007). Effective PD efforts help teachers align the content standards that they are responsible for teaching with the practical delivery and implications of classroom implementation.

Second, effective PD should be content focused. Desimone and colleagues (2002) conducted a longitudinal survey study, over 3 years, of 207 teachers in 30 schools. Teachers reported that PD focused on specific instructional practices increased their use of that instructional practice. Participant responses to the surveys also indicated that, in creating content focused PD, it is important to carefully blend theory, content area knowledge, understanding of students as learners, and general pedagogical skills. These elements help ensure that teachers not only understand the content, but also are able to properly implement content area knowledge that connects to student needs. In a mixed method case study, Vavasseur and MacGregor (2008) investigated how the professional development of middle school teachers was facilitated through their participation in content-focused online communities of practice. Results indicated that, in this content-focused context, teachers gained curriculum-based knowledge, developed enhanced self-efficacy with respect to implementing technology, and collaborated on the development of interdisciplinary curriculum units.

Third, effective PD includes monitoring of student gains. Faulkner and Cain (2013) conducted a 5-day professional development module to improve teachers' math knowledge and understanding of number sense. Using a quasi-experimental design, they investigated if the students enrolled in the trained teachers' classrooms improved in mathematics content

performance. Significant changes in student performance were not found, which was attributed to teachers not continuously monitoring student achievement and student responses. Because this may be difficult while also learning new instructional material or interventions, an element of effective PD may be to provide teachers with tools to monitor gains in student achievement (Yoon, Duncan, Lee, & Shapley, 2008).

Fourth, it is suggested that effective PD is active. Researchers suggest that teachers be actively engaged in learning new instructional practices, preferably in actual classroom settings (Dagan & Bean, 2014; Desimone et al., 2002; McCutchen et al., 2002). Browder and colleagues (2012) developed an interactive PD on alternate achievement standards for 193 teachers of students with severe disabilities. Results of this quasi-experimental study indicated that teachers improved in aligning standards and generalizing the training to other content areas. It is suggested that teachers receiving PD obtain tangible images and demonstrations on how to apply new instructional strategies in the classroom (Bryant, Linan-Thonpson, Ugel, Hamff, & Hogan, 2001).

Finally, effective PD can be collaborative in nature (Brock & Carter, 2013; Garet et al., 2001; Penuel et al., 2007). Teachers may benefit from continuous feedback, opportunities to observe, and coaching inside the classroom when implementing new materials (Billingsley, 2005; Brownell, Ross, Colon, & McCallum 2005; Neufeld & Roper, 2003). Researchers have suggested that during and after PD, teachers and school personnel should work together to share ideas, discuss problems that may arise, discuss student response to new material, and share instructional resources needed for proper implementation (Wayne, Yoon, Zhu, Cronen, & Garet, 2008).

Reform PD

These elements of effective PD move the field away from more traditional types of PD toward what has been referred to as *reform PD* (Darling-Hammond & McLaughlin, 1995). Penuel and colleagues (2007) conducted a survey of 454 teachers involved in an inquiry science program to examine the effects of different characteristics of PD. This study points to the significance of teachers' perceptions about how coherent their professional development experiences were for teacher learning and program implementation. Teachers reported that the following aspects of reform PD increased their contact with the content being acquired: (1) teacher study groups, (2) teacher networks or communities, (3) mentoring or coaching, (4) internships, and (5) resource centers. It has been suggested that reform PD is more effective than traditional PD because it requires that teachers take time to explore new theories and strategies, introduced briefly, to improve classroom practice. According to Hargreaves (2014), of the reform-oriented PD elements, teacher coaching is the most widely used and has been mentioned in professional literature as an effective means to encourage necessary follow-up and on-going discussion about professional learning.

Coaching. Defined as the study and teaching of theory, the observation of demonstrations, and opportunities for feedback given by a peer, mentor, supervisor, or PD provider (Showers, 1982), coaching is a critical support for persistent use of newly adopted practices (Joyce & Showers, 1982). According to Fixsen and colleagues (2005), coaching is a core component of effective program implementation because newly acquired behavior is (a) rudimentary when compared to

routine of a more experienced practitioner, (b) delicate and needs to be reinforced for consumers in a natural setting, and (c) imperfect and will need to be formed to be most functional in a natural setting. The form coaching takes differs in the literature, but it is likely to include some combination of the following components: planning, teaching, modeling, practicing new skills, direct supervision of implementation of target practices in the classroom, and feedback (Fixsen et al., 2005; Stormont & Reinke, 2012). A growing body of research substantiates the positive effects coaching can have on both teacher behavior and student outcomes. For example, Bradshaw and colleagues (2008) conducted a quasi-experimental study on using PD and coaching to implement school-wide positive behavior intervention supports (SWPBIS). Twenty-one schools received training, and 16 schools did not. The schools receiving PD and coaching showed significant gains in SWPBIS implementation as compared to the control group.

Supervisory Teacher Coaching. An outside expert or supervisor who gives specific, positive, and corrective coaching when needed provides supervisory teacher coaching. This type of coaching can be offered to the teacher after the completion of the observed lesson or in an effort to move the recipient toward a desired level of performance (Joyce & Showers, 1981; 1982; Maeda, 2001; Simonsen, Myers, & DeLuca, 2010). Supervisory teacher coaching offers a supportive means for teachers to implement new learning and improve fidelity over time (Joyce & Showers, 1982; Miller, Harris, &Watanabe, 1991; Shidler, 2009; Showers 1985). Supervisory coaching offers an extension of PD from an expert or supervisor that offers an in-depth study of the theory behind the skill being coached, observations of demonstrations, and continuous practice with timely feedback.

Supervisory teacher coaching as a reform PD activity has a growing evidence base; however, the relationship between supervisory coaching and positive teacher behaviors remains unclear. That is, in some studies, supervisory coaching results in significant change in teacher behavior, while in other studies no significant effects are observed. Moreover, the quality of these studies is inconsistent, with some designed according to standards for conducting impactful, high-quality research and others only meeting some of those standards. Therefore, the purpose of this systematic review of the literature is to answer the following questions:

1. What studies can be identified as implementing supervisory coaching?
2. Among these studies, what are the characteristics and components of supervisory teacher coaching that lead to successful teacher and student outcomes?
3. What is the quality of the existing literature on supervisory teacher coaching?

Method

Search and Selection Process

A search was conducted using the following EBSCO databases: Academic Search Complete, Education Full Text, Professional Development Collection, PsycARTICLES, PsycEXTRA, Psychology and Behavioral Sciences Collection, PsycINFO, and Sociological Collection. An additional search was conducted using ProQuest and Google Scholar. The search terms used included the following: *special education and teacher coaching and k-12 education and teacher coaching* or *supervisory coaching* or *in vivo coaching* and *professional development* or *staff development* or *teacher training* or *in-service* and *teacher* or *practitioner* or *educator* or *instructor*.

To be included in this systematic review of literature, studies had to meet six inclusionary criteria: (a) conducted with special or general education teachers; (b) conducted with teachers of k-12 students; (c) utilized a supervisory coaching model as the independent variable in the study; (d) published in a peer reviewed journal; (e) use of experimental or quasi-experimental research design to show a causal relationship between supervisory teacher coaching and teacher behavior; (f) conducted in the United States. As a result of these requirements, articles containing the following were excluded: (a) teachers certified in bilingual education subjects only; (b) settings in preschool classrooms; (c) utilization of peer coaching models; (d) use of qualitative methods; and (e) studies conducted in classrooms outside the United States. An ancestral search was conducted on all articles that met criteria. The search originally rendered 277 articles, 16 of which met criteria for inclusion (see Table 1).

Quality Indicators

Intervention research sets out to establish that the desired change in participants' behavior was caused by the intervention and not because of any other reasonable explanation. To assess the quality of experimental and quasi-experimental, Gersten and colleagues (2005) set forth a set of quality standards. Gersten et al. provided 10 quality indicators within four domains for evaluating experimental and quasi-experimental research (see Table 2). The six experimental and quasi-experimental design studies were coded using Gersten et al.'s (2005) standards and the following guiding questions:

1. *Participants and Settings* addresses three questions: (a) are participant difficulties or disabilities described sufficiently?; (b) are relevant characteristics comparable across conditions?; and (c) are characteristics of interventionist/teachers described sufficiently, and are they comparable across conditions?
2. *Implementation of Intervention and Description of Comparison Conditions* addresses three questions: (a) are interventions clearly and specifically described?; (b) is fidelity of implementation described and assessed?; and (c) are comparison conditions described?
3. *Outcome Measures* addresses two questions: (a) are multiple measures closely aligned with intervention?; and (b) are outcomes for capturing the intervention's effect measured at appropriate times?
4. *Data Analysis* addresses two questions: (a) are techniques linked to research questions?; and (b) did the research report include inferential statistics and effect size calculations?

To systematically assess the quality of single-subject research, Horner and colleagues (2005) provided 18 quality indicators within seven domains for evaluating single-subject studies (see Table 3). The ten single case design studies were coded using Horner et al.'s (2005) standards and the following guiding questions:

1. *Participants and Setting* addresses three questions: (a) are the participants described sufficiently to be replicated by the reader?; (b) is the selection of participants described sufficiently to be replicated by the reader?; and (c) is the setting of the study described sufficiently to be replicated by the reader?
2. *Dependent Variable* addresses five questions: (a) was it described well enough to be replicated?; (b) was it observable?; (c) how it was measured?; (d) how often it was

measured?; and (e) was inter-observer reliability reported, and did it meet minimum levels of acceptability (e.g. IOA= 80%)?

3. *Independent Variable* addresses three questions: (a) was it described sufficiently enough to be replicated?; (b) was it systematically manipulated?; and (c) was procedural fidelity described and measured?
4. *Baseline* addresses two questions: (a) was the condition of baseline described sufficiently as to be replicated by the reader?; and (b) was evidence provided regarding baseline patterns and trends?
5. *Experimental Control/Internal Validity* addresses three questions: (a) were there three demonstrations of experimental effect?; (b) did the design control for threats to internal validity?; and (c) did the results indicate a pattern that demonstrated experimental control, as judged by visual analysis?

Procedures

Studies chosen to be included in this review of the literature were evaluated to identify student and teacher participants, settings, dependent variables, independent variables, design, and results. To analyze the chosen studies, each domain was rated *Yes* if all quality indicators were present. If one quality indicator was not present in a domain, that domain was rated *No*. In addition, the quality indicator that dealt with procedural fidelity only received a *Yes* rating if fidelity had been collected on the actual coaching procedures. Finally, studies were analyzed for components of coaching procedures commonly identified in the literature: (a) modeling, (b) prior training, (c) pre-observation meeting, (d) e-mail feedback, (e) handwritten feedback, (f) face-to-face feedback, and (g) teacher goal setting.

Reliability was coded for the evaluation of the quality of the articles after the literature review was completed. An independent coder was trained on the established inclusion criteria and search methods, including a mock search to agree on search terms and criteria. Researcher and independent coder reached 100% agreement during training. Then, the independent coder read all chosen articles and reviewed them for basic evaluation criteria, quality indicators and components of coaching procedures. All total, 588 items were coded. Each research question was coded: research evaluation (n = 112), coaching components (n = 112), and quasi-experimental (n = 84) and single case (n = 280) quality indicators.

The formula for calculating reliability involved dividing the number of agreements by the sum of agreements and disagreements and multiplying by 100. Reliability agreement was 100% for the total number of articles. Reliability on the research evaluation reached 90.1%. Reliability on the information coded for the quasi-experimental studies was 98.8%. Reliability on single case studies was 97.5%, and the reliability on coaching components evaluated was 97.3%.

Results

Supervisory Coaching Study Evaluation

A total of 485 teacher participants were included in the studies that met the criteria for this review of literature. Of those teachers, 11 (3%) were teachers in gifted/honors classrooms, 76 (15%) were classified as in-service special education teachers, 292 (60%) were general education teachers, and 106 (22%) were dually certified as general education and English language learner

(ELL) teachers. One teacher was a preschool teacher who did not meet inclusion criteria; therefore, that teacher's data were not included in the analyses. The teachers included in these studies taught students in a variety of settings. While 75% of the studies were conducted in an elementary classroom, studies also were conducted in alternative schools (17.65%), middle schools (23.5%), and high schools (17.65%). The majority of the studies (56.25%) included in this review were conducted in a general education classroom; 37.5% of studies were conducted in a self-contained special education classroom; and 6.25% of studies were conducted in an inclusion classroom.

Over 56% of the studies discussed in this review of literature used teacher praise as a teacher participant dependent variable. That is, a coaching intervention was used to increase the use of teacher praise. Of the remaining studies, 31.25% (n = 5) measured the teachers' fidelity of implementation of the function-based intervention; fidelity was maintained after the coaching intervention was removed for two weeks. Kretlow et al. (2011; 2012) observed an increase in accuracy of delivery of academic tasks, i.e., math problem solving, use of response cards, by teachers with the introduction of teacher coaching. An immediate increase in desired academic tasks was observed in both studies with the onset of supervisory coaching. The improved fidelity of implementation of an established behavior plan was measured in Codding and colleagues (2005). Teachers received a coaching intervention, and with the onset of intervention, fidelity of implementation improved and was maintained even after the end of the intervention phase.

In this review, supervisory teacher coaching was used as a stand-alone intervention in 23.5% of the studies. The remaining studies offer supervisory coaching along with PD. Of the studies reviewed, general education teachers displayed relatively positive results with some change in teacher behaviors after the introduction of supervisory coaching. Although studies focusing on special education classrooms also displayed positive results, special education teachers represented only small number of participants (Gregory et al., 2013; Kretlow et al., 2011; Kretlow et al., 2012; Matsumara et al., 2012, Podhajski et al., 2009; Rienke et al., 2014, Sailors & Price, 2015; Thompson et al., 2012).

Components of Coaching

Fidelity of implementation on the coaching procedure was examined, as well as the description of components of supervisory coaching, in each study (see Table 4). Fidelity of implementation was collected on coaching procedures in only 50% of examined studies. Face–to-face feedback was offered in 94% of studies. Components offered along with face-to-face coaching included email feedback (12%) and face-to-face coaching offered via web (12%). Of the studies reviewed, only two studies explored the option of using widely available web-based technologies. Ruble et al. (2013) compared web-based coaching to face-to-face coaching with similar gains in positive teacher behaviors, compared to the placebo group that received no coaching. Gregory et al. (2013) showed a modest increase in student engagement with the use of the "My Teaching Partner" web-based program in secondary classrooms. Training was offered to participants in 81% of studies, while modeling and pre-observation meetings were offered in 50% of studies. Handwritten feedback was offered in 44% of studies, while teacher goal setting was utilized in only 31% of studies. Location of coaching was discussed in 31%, and time spent coaching, or dosage, was mentioned in 37.5% of reviewed studies.

Evaluation of Research Quality

Of the six experimental or group design studies, two (Ruble et al., 2009; Sailors & Price, 2015) met all four quality indicator domains for group design studies. The remaining group design studies failed to describe or report the collection of treatment fidelity, therefore not meeting one domain of quality indicators.

Of the ten single case design studies, only four (Bethune & Wood, 2013; Duchaine et al., 2011; Kretlow et al., 2012; Kretlow et al., 2011) met all quality indicator domains and 3 of the 4 studies measured the fidelity of implementation of a specific academic task. The remaining single case studies failed to meet indicators in the dependent variable, independent variable, baseline, internal validity, and social validity domains. Although the studies did meet most of the domains, each domain and sub-domain had to have been met (*Yes*) for the study to meet all quality indicators.

Of the single case studies, Marten et al. (1997) utilized a multiple baseline design with only two cases as opposed to the three cases suggested by What Works Clearinghouse (Kratochwill et al. (2010), therefore not meeting requirements for internal validity quality indicators. Sutherland and colleagues (2000) used a withdrawal (ABAB) single case design to measure an intervention that is seemingly not reversible. Sutherland collected baseline data, introduced coaching, and withdrew the coaching intervention. Kazdin (2011) would suggest that the participant could not withdraw information that has been learned. The remaining single case studies failed to meet indicators in the dependent variable, independent variable, baseline, internal validity, and social validity domains. Single case studies also did not meet indicators in the dependent variable, independent variable, baseline, internal validity, and social validity domains (Codding et al., 2005; Martens et al., 1997; Miller et al., 1991; Simonsen et al., 2010; Sutherland et al., 2000; Thompson et al., 2012).

Discussion

The purpose of this systematic review of the literature was to examine the literature that tested the effects of supervisory teacher coaching with K-12 classroom teachers. Results of this literature review suggest that supervisory teacher coaching may be a promising strategy to change classroom teacher behaviors. The results have been mixed. In general, supervisory teacher coaching increased desired behaviors in classroom teachers; however, the behavior was not always maintained nor was maintenance always measured. Duchaine and colleagues displayed mixed results among teachers with the maintenance of behavior specific praise statements. A majority of studies included in this review did not collect maintenance data at all (Gregory et al.,2014; Kretlow et al., 2012; 2011; Martens et al., 1997; Matsumara et al., 2010; Podhajski et al., 2009; Reinke et al., 2014; Ruble et al., 2013; Sailors et al., 2015; Simonsen et al., 2010; Sutherland et al., 2000; Thompson et al., 2012) therefore it is not possible to know for sure if the teachers effectively maintained the strategy introduced by the coach. Student data were also mixed. In studies where student data were collected, they did not always show completely positive results. Where student data were collected students often made small gains or gains that were not maintained without the intervention. The included studies seem to assert that supervisory teacher coaching has a more positive effect on teacher outcomes than student outcomes.

Over half of the studies reviewed concentrated on coaching to increase teacher praise. In each study, teacher praise increased with the support of the coaching intervention. In addition to teacher praise, one study found an increase in the use of higher level questioning and a decrease in general, nonspecific praise. Although supervisory coaching did not have the same positive outcomes on teacher-specific behaviors as it did on students' academic performance, it is important to note that the teacher-specific behaviors are of interest in these studies. Studies that focused on academic outcomes showed gains in student performance and teacher implementation of the specific academic task.

Another important finding from the review was that several specific components of supervisory coaching are often not described in the literature. These components are necessary to evaluate the effectiveness of a coaching package and to replicate implementation of the intervention. It is also an important factor in informing future research in coaching and creating a solid evidence base for the field. Details such as location and dosage of coaching could have a substantial impact on the outcome.

Finally, when standards for high quality research were applied to the reviewed studies, the results indicated that quality was highest in both single case and quasi-experimental design studies where teachers received coaching to improve the fidelity of implementation of specific academic tasks (e.g., reading comprehension, calendar math), as compared to when it was introduced to change teacher-specific behaviors (e.g., praise statements, opportunities to respond). Not only did studies concentrating on specific academic tasks meet more quality indicators (Gersten et al., 2005; Horner et al., 2005), they also demonstrated more immediate and sustained positive outcomes (Bethune & Wood, 2013; Kretlow et al., 2011; Kretlow et al., 2012).

The discrepancy in the adherence to quality indicators could be connected to the fact that it may be easier to measure data on a specific academic task as opposed to teacher-specific behaviors where guidelines may not be as clear as in specific academic tasks. Of the majority of studies that offered face to face coaching, only Bethune and Wood, 2013; Duchaine et al., 2011; Kretlow et al., 2012; Kretlow et al., 2011; Sailors & Price, 2015; Ruble et al, 2009 met all quality indicators.

Although we attempted to do a comprehensive review of the current evidence base surround supervisory coaching in k-12 classrooms, there is a possibility that studies that should have been included were not found. There may have been other search terms we should have added to our list or studies missed in manual ancestral searches.

Implications for Future Research

This systematic review of the literature revealed gaps in the teacher coaching literature that should be addressed in future research. First, all studies included in this review were conducted using university researchers or coaches trained by university researchers to implement the intervention. Future research could explore the effectiveness of school personnel providing the coaching (e.g., academic coaches, school psychologists, administrators). Further investigation would be necessary to determine whether the supervisory teacher coaching interventions would generalize with k-12 classroom teachers using school personnel as the coach of record.

Second, follow-up or maintenance data may be collected to add information to the literature base on the sustainability of the evidence based practice that the teacher was coached to use. Collecting follow-up data is important in these investigations, as it provides information on the extent that continuous use of teacher coaching is needed. Third, it is suggested that future researchers ensure that quality indicators are adhered to in an effort to increase the output of quality research that is replicable and reliable. For researchers interested in delving into the effects of specific teacher behaviors, such as specific praise or opportunities to respond, it is recommended that researchers maintain strict standards and that those standards be properly and thoroughly explained to ensure valuable research based on proposed quality indicators.

Fourth, in planning and carrying out coaching studies, it is recommended that components of the coaching package being used (e.g. coaching dosage, location of coaching) are carefully described. The components need to be described not only to increase the likelihood of accurate replication but also to ensure that researchers can identify which components have the greatest effect on the intervention outcome. Fidelity of implementation also should be collected to ensure that the coaching procedure is implemented as described.

Fifth, it is suggested that future researchers work to expand research into special education and secondary classrooms. The number of special education teachers and students included in published coaching studies is very low compared to the larger number of general education teachers and students. Moreover, of the special education classrooms included in this review, the majority of study participants were teachers in self-contained classrooms with students who have severe disabilities. Teachers of students with high-incidence disabilities and in collaboratively taught classrooms only represented a small number of included studies. Another disparity in population was the very low number of secondary classrooms included. Teachers in secondary classrooms could benefit from supervisory coaching on newly acquired academic tasks learned in PD and on implementing behavioral interventions for students who display challenging behaviors. Going forward, researchers may concentrate on those populations and classroom settings, which may support both generalizability and impact of this approach to coaching. Finally, researchers should continue to explore the viability of technology-based coaching for teachers in k-12 classrooms. As technology continues to be more immersed in aspects of daily life, researchers should explore the option of offering PD and follow-up coaching via technology. Using technology to train teachers may not only be a cost effective option; it also may offer flexibility for the coach and teacher.

Conclusion

This review revealed that desirable teacher behaviors can improve with the use of supervisory teacher coaching in several settings. This review indicated that supervisory teacher coaching might be a practical strategy for improving and increasing teachers' use of evidence-based strategies. While the studies included in this review displayed some positive change in teacher behavior, it is not possible to say conclusively that the supervisory teacher coaching intervention will have a long term positive impact on teacher behaviors without additional research.

References

Bethune, K. S., & Wood, C. L. (2013). Effects of coaching on teachers' use of function-based interventions for students with severe disabilities. *Teacher Education and Special Education: The Journal of the Teacher Education Division of the Council for Exceptional Children*, *36*, 97-114.

Billingsley, B. S. (2005). *Cultivating and keeping committed special education teachers: What principals and district leaders can do*. New York: Corwin Press.

Birman, B. F., Desimone, L., Porter, A. C., & Garet, M. S. (2000). Designing professional development that works. *Educational Leadership*, *57*(8), 28-33.

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher*, *33*(8), 3-15.

Bradshaw, C. P., Reinke, W. M., Brown, L. D., Bevans, K. B., & Leaf, P. J. (2008). Implementation of school-wide positive behavioral interventions and supports (PBIS) in elementary schools: Observations from a randomized trial. *Education and Treatment of Children, 23* (4) 1-26.

Brock, M. E., & Carter, E. W. (2015). Effects of a professional development package to prepare special education paraprofessionals to implement evidence-based practice. *Journal of Special Education*, *49*(1), 39-51.

Browder, D. M., Jimenez, B., Mims, P. J., Knight, V. F., Spooner, F., Lee, A., & Flowers, C. (2012). The effects of a "Tell-Show-Try-Apply" professional development package on teachers of students with severe developmental disabilities. *Teacher Education and Special Education*, *35*, 212-227.

Brownell, M. T., Ross, D. D., Colon, E. P., & McCallum, C. L. (2005). Critical features of special education teacher preparation: A comparison with general teacher education. *The Journal of Special Education*, *38*, 242-252.

Bryant, D. P., Linan-Thompson, S., Ugel, N., Hamff, A., & Hougen, M. (2001). The effects of professional development for middle school general and special education teachers on implementation of reading strategies in inclusive content area classes. *Learning Disability Quarterly*, *24*, 251-264.

Codding, R. S., Feinberg, A. B., Dunn, E. K., & Pace, G. M. (2005). Effects of immediate performance feedback on implementation of behavior support plans. *Journal of Applied Behavior Analysis*, *38*, 205-219.

Cullinan, D., & Sabornie, E. J. (2004). Characteristics of emotional disturbance in middle and high school students. *Journal of Emotional and Behavioral Disorders*, *12*(3), 157-167.

Dagan, A. & Bean, R. (2014). High quality research based professional development. S. Kragler (Ed.), *Handbook of Professional Development in Education: Successful Models and Practices, PreK-12*. (pp. 42-64) New York: Guilford Publications.

Darling-Hammond, L. (2014). Strengthening clinical preparation: The holy grail of teacher education. *Peabody Journal of Education*, *89*(4), 547-561.

Darling-Hammond, L., & McLaughlin, M. W. (1995). Policies that support professional development in an era of reform. *Phi delta kappan*, *76*(8), 597-604.

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. *Educational Evaluation and Policy Analysis*, *24*, 81-112.

Duchaine, E. L., Jolivette, K., & Fredrick, L. D. (2011). The effect of teacher coaching with performance feedback on behavior-specific praise in inclusion classrooms. *Education and Treatment of Children*, *34*, 209-227.

Elish-Piper, L., & L'Allier, S. K. (2010). Exploring the relationship between literacy coaching and student reading achievement in grades K–1. *Literacy Research and Instruction*, *49*(2), 162-174.

Faulkner, V. N., & Cain, C. R. (2013). Improving the mathematical content knowledge of general and special educators evaluating a professional development module that focuses on number sense. *Teacher Education and Special Education*, *36*(2), 115-131.

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature.* Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network.

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, *38*, 915-945.

Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children*, *71*, 149-164.

Gregory, A., Allen, J. P., Mikami, A. Y., Hafen, C. A., & Pianta, R. C. (2014). Effects of a professional development program on behavioral engagement of students in middle and high school. *Psychology in the Schools*, *51*, 143-163.

Hargreaves, A. (2014). Measuring effectiveness of educators' professional development. L. E. Martin, S. Kragler, D. J. Quatroche, & K. L. Bauserman (Eds.) *Handbook of Professional Development in Education: Successful Models and Practices, PreK-12* (pp. 447-466) New York: Guilford Publications.

Haydon, T., Conroy, M. A., Scott, T. M., Sindelar, P. T., Barber, B. R., & Orlando, A. M. (2010). A comparison of three types of opportunities to respond on student academic and social behaviors. *Journal of Emotional and Behavioral Disorders, 18*(1), 27-40.

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children*, *71*, 165-179.

Joyce, B. R., & Showers, B. (1981). Transfer of training: The contribution of "coaching". *Journal of Education*, 163-172.

Joyce, B., & Showers, B. (1982). The coaching of teaching. *Educational leadership*, *40*(1), 4-10.

Klingner, J. K. (2004). The science of professional development. *Journal of Learning Disabilities*, *37*, 248-255.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). Single-case designs technical documentation. *What Works Clearinghouse*.

Kretlow, A. G., Cooke, N. L., & Wood, C. L. (2012). Using in-service and coaching to increase teachers' accurate use of research-based strategies. *Remedial and Special Education*, *33*, 348-361.

Kretlow, A. G., Wood, C. L., & Cooke, N. L. (2011). Using in-service and coaching to increase kindergarten teachers' accurate delivery of group instructional units. *Journal of Special Education*, *44*, 234-246.

Leko, M. M., & Brownell, M. T. (2009). Crafting quality professional development for special educators: What school leaders should know. *Teaching exceptional children*, *42*, 64-70.
Loucks-Horsley, S. (1998). The role of teaching and learning in systemic reform: A focus on professional development. *Science Educator*, *7*, 1-6.
Maeda, J. K. (2001). Teacher coaching in physical education: A review. *Physical Educator*, *58*(3), 140.
Martens, B. K., Hiralall, A. S., & Bradley, T. A. (1997). A note to teacher: Improving student behavior through goal setting and feedback. *School Psychology Quarterly*, *12*, 33.
Matsumura, L. C., Garnier, H. E., Correnti, R., Junker, B., & Bickel, D. D. (2010). Investigating the effectiveness of a comprehensive literacy coaching program in schools with high teacher mobility. *The Elementary School Journal*, *111*, 35-62.
McCutchen, D., Abbott, R. D., Green, L. B., Beretvas, S. N., Cox, S., Potter, N. S., & Gray, A. L. (2002). Beginning literacy links among teacher knowledge, teacher practice, and student learning. *Journal of Learning Disabilities*, *35*, 69-86.
Merriam, S. B., Caffarella, R. S., & Baumgartner, L. M. (2012). *Learning in adulthood: A comprehensive guide*. San Francisco: John Wiley & Sons.
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: The Journal of the Teacher Education Division of the Council for Exceptional Children, 14*, 183-191.
National Statistics for Education Statistics. (2007). The condition of education 2007 (NCES 2007-064). Washington, DC: US. Department of Education. Retrieved from http://nces.ed.gov/pubs2007/2007064.pdf
Neufeld, B., & Roper, D. (2003). *Coaching: A strategy for developing instructional capacity: Promises & Practicalities*. Aspen Institute.
Penuel, W. R., Fishman, B. J., Yamaguchi, R., & Gallagher, L. P. (2007). What makes professional development effective? Strategies that foster curriculum implementation. *American Educational Research Journal*, *44*, 921-958.
Podhajski, B., Mather, N., Nathan, J., & Sammons, J. (2009). Professional development in scientifically based reading instruction teacher knowledge and reading outcomes. *Journal of Learning Disabilities*, *42*, 403-417.
Putnam, R. T., & Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teacher learning? *Educational Researcher*, *29* (1) 4-15.
Reinke, W. M., Stormont, M., Herman, K. C., Wang, Z., Newcomer, L., & King, K. (2014). Use of coaching and behavior support planning for students with disruptive behavior within a universal classroom management program. *Journal of Emotional and Behavioral Disorders, 22*(2), 74-82.
Ruble, L. A., McGrew, J. H., Toland, M. D., Dalrymple, N. J., & Jung, L. A. (2013). A randomized controlled trial of COMPASS web-based and face-to-face teacher coaching in autism. *Journal of Consulting and Clinical Psychology*, *81*, 566.
Sailors, M., & Price, L. (2015). Support for the improvement of practices through intensive coaching (SIPIC): A model of coaching for improving reading instruction and reading achievement. *Teaching and Teacher Education*, *45*, 115-127.
Simonsen, B., Myers, D., & DeLuca, C. (2010). Teaching teachers to use prompts, opportunities to respond, and specific praise. *Teacher Education and Special Education: The Journal of the Teacher Education Division of the Council for Exceptional Children, 33*, 300-318.

Shidler, L. (2009). The impact of time spent coaching for teacher efficacy on student achievement. *Early Childhood Education Journal*, *36*(5), 453-460.

Showers, B. (1982). Transfer of Training: The Contribution of Coaching.

Showers, B. (1985). Teachers coaching teachers. *Educational leadership*, *42*(7), 43-48.

Stormont, M., & Reinke, W. M. (2012). Using coaching to support classroom-level adoption and use of interventions within school-wide positive behavioral interventions and support systems. *Beyond Behavior*, *21*(2), 11-19.

Sutherland, K. S., Wehby, J. H., & Copeland, S. R. (2000). Effect of varying rates of behavior-specific praise on the on-task behavior of students with EBD. *Journal of Emotional and Behavioral Disorders*, *8*, 2-8.

Thompson, M. T., Marchant, M., Anderson, D., Prater, M. A., & Gibb, G. (2012). Effects of tiered training on general educators' use of specific praise. *Education and Treatment of Children*, *35*, 521-546.

Vavasseur, C. B., & Kim MacGregor, S. (2008). Extending content-focused professional development through online communities of practice. *Journal of Research on Technology in Education*, *40*(4), 517-536.

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*, 469-479.

Yoon, K. S., Duncan, T., Lee, S. W. Y., Scarloss, B., & Shapley, K. L. (2007). Reviewing the evidence on how teacher professional development affects student achievement. Issues & Answers. REL 2007-No. 033. *Regional Educational Laboratory Southwest (NJ1).*

Table 1

Overview of Supervisory Teacher Coaching Studies

Citation	**Teacher Participant**	**Student Participant**	**Setting**	**DV**	**IV**	**Design**	**Results**
Bethune & Wood, (2013)	3 special education teachers, 4yrs exp M, 3yrs exp, 15yrs exp, 2 female, 1 male, SpEd Cert, Did not report age	4 Students w/ moderate to severe disabilities w/challenging BX, 1-DS – 8 yo, 1-SevAutism – 5yo, 1 – autism-10yo	Elementary school self-contained	Percent accuracy of teacher implementation of function based intervention -Measure of the students problem behavior - Measure of the students replacement behavior	Researcher implemented side by side coaching	Delayed MB across participants design for teachers, MB across participants design for students	Teacher implementation function based interventions improved. Student problem behavior decreased with FBI and coaching. Replacement bx results varied.
Codding et al. (2005)	5 special education teachers, Experience ranged from, 6-30 months,	3 students with non-traumatic brain injury, -2 with TBI, Ages 10-19	Private school for students with brain injuries ages 10-19	Integrity of steps of behavior plan implemented as written.	Face to face Performance feedback	Concurrent MB across teacher-student dyads	Increase with all 5 teacher-student dyads on proper implementation of antecedent

Citation	**Teacher Participant**	**Student Participant**	**Setting**	**DV**	**IV**	**Design**	**Results**
	Bachelor's degrees, SpEd Cert, Did not report age						and consequence components. The results varied; all of the teachers did not respond to the entire intervention.
Duchaine et al. (2011)	1 sp ed, - MA degree 2 gen ed – 1 BA 1 MA	Random sample of students with and w/out disabilities	High School Inclusion math	Teacher Praise, On-task BX	Written performanc e feedback with training	Multiple Baseline across teachers	Increase in Praise. Inconclusive w/ on-task BX; Maintenance showed return to baseline BX.
Gregory et al. (2013)	87 teachers in 12 schools	n/a	61% middle school general ed 39% high school	student engagement	My Teaching Partner - Secondary	Random ized control trial	Although, modest increase in student bx engagement. Results were mixed; the

Citation	**Teacher Participant**	**Student Participant**	**Setting**	**DV**	**IV**	**Design**	**Results**
							Emotional Support domain of the CLASS-S the intervention did not result in a major shift
Kretlow et al. (2012)	3 1st grade teachers	28-30 students per class 84-90 students total	Elementary school 1st grade classroom	Accuracy in implementation of trained calendar math and an untrained math practice	Professional Development and coaching	Multiple baseline across teachers	Mean instructional accuracy increased post in-service and again post coaching. No student data collected
Kretlow et al. (2011)	3 kindergarten teachers	n/a	Elementary school kindergarten classroom	Accuracy in group instructional units	Professional Development and coaching	multiple baseline	Mean instructional accuracy increased post in-service and again post coaching. No causal relation in-

Citation	**Teacher Participant**	**Student Participant**	**Setting**	**DV**	**IV**	**Design**	**Results**
							service training was simultaneous .
Martens et al. (1997)	1 special education teacher, Female, SpEd Cert, Did not report age or degree	2 boys w/ ED, 6 yrs old	Self-contained elementary classroom	Combined appropriate BX (schoolwork , attending to instruction, responding aloud)	Goal setting plus written feedback	Multiple baseline across students design	Appropriate BX increased immediately and remained stable with intervention. Study does not meet evidence based practice standards
Matsumara et al. (2010)	177 4th and 5th grade teachers (79 left) - 2nd year 171(73 added) 4th and 5th grade teachers	1269 student standardized student test scores	32 Elementary schools	state standardized test for English language learners	Content focused coaching	HLM analyses	CFC program predicted significantly higher school level gains on the state standardized test for English language learners.

Citation	**Teacher Participant**	**Student Participant**	**Setting**	**DV**	**IV**	**Design**	**Results**
							Only half the intended teachers participated in the study.
Miller et al. (1991)	6 special education teachers , 2 M, 4 B, SpEd Cert, Did not report age or sex	N/A	Elementary, Middle & High summer school for low achievers	Effective and ineffective teacher BX	Florida Performance Measurement System (coaching form)	Multiple baseline across teachers	Effective teacher BX increased while ineffective practices remained low. Results were mixed, after coaching, neither teacher in team 3 demonstrated improvements in teaching behaviors. No Student data collected.

Citation	**Teacher Participant**	**Student Participant**	**Setting**	**DV**	**IV**	**Design**	**Results**
Podhajski et al. (2009)	Experimental group 4 1st and 2nd grade teachers Control group 3 1st and 2nd grade teachers	Experimental - 33 1st grd 20 2nd grade students - 3 504 3 IEPs Control - 14 1st grd, 22 2nd grd - 1 504, 1 IEP	Public school 1st and 2nd grade classrooms	scientifically based reading instruction	TIME for teachers professional development plus coaching	Randomized control trial	Teacher growth on teacher knowledge and student growth in student measures. Mixed results; TOWRE Sight Word Efficiency showed no growth pretest to posttest.
Reinke et al. (2014)	68 teachers 34 teachers received coaching	1,148 total students	Grades k-3 grades	Teacher reprimands, teacher use of general and specific praise	Coaching using universal classroom management program	quasi-experimental	Decrease in teacher reprimands. Although low, increase in teacher general and specific praise

Citation	**Teacher Participant**	**Student Participant**	**Setting**	**DV**	**IV**	**Design**	**Results**
Ruble et al. (2013)	49 special education teachers, 1 male , 48 female, SpEd Cert, 22 – B 45%, 23 – M 47%, 4 – DNI 8%, Did not report age	49 students with Autism (1 per teacher) ages 3-9 years old	Elementary, one to one work on IEP goals	IEP goals as measured with PET-GAS tool	COMPASS coaching tool	Randomized control trial	COMPASS tool improves IEP goal outcomes for students with Autism than the placebo group. Unable to detect differences in web group versus face-to-face group.
Sailors & Price (2015)	120 teacher - 50 control - 70 treatment - grades 2(16%), 3(12%), 4(13%), 5(15%), 6(17%), 7(15%), 8(12%)	1496 student participants	3 school districts in Texas 2-8th grade classrooms	comprehension instruction	workshops plus SIPIC (Support for the Improvement of Practices through Intensive Coaching) coaching model	quasi-experimental - pre/post test	Coached teachers improved comprehension instruction. Also, SIPIC model of coaching improved the practices of reading teachers and

Citation	**Teacher Participant**	**Student Participant**	**Setting**	**DV**	**IV**	**Design**	**Results**
							increased the student achievement.
Simonsen et al. (2010)	3 special education teachers , 2 female, 1 male, SpEd Cert, M 16yr exp, 13 yr exp, M 13 yrs exp, Did not report age	N/A	Self-contained classrooms alt setting	Prompts, OTRs and SP	PORT training and feedback	Multiple baseline across DV	Training alone did not increase praise or OTR. Feedback increased praise & OTR.
Sutherland et al. (2000)	1 special education teacher – male Did not report age or degree	9 students with E/BD, 2 girls, 7 boys, ages 10-11, 6 Black, 3 White	5th grade self-contained classroom	Non-behavior specific praise - Behavior specific praise -On-task behavior	Observation and verbal feedback	ABAB withdrawal design	NBSP and BSP increased with observer feedback. On-task BX improved with the increase of NBSP & BSP. No

Citation	Teacher Participant	Student Participant	Setting	DV	IV	Design	Results
							student data collected; replication during academic instruction is needed to determine on-task behavior of students.
Thompson et al. (2012)	3 Female teachers between ages 40-50	3 students ages 8,10&11 displayed non-compliant/disruptive bx	3 public elementary schools	Teacher BSP - Student On-Task	Training, Video Self-Monitoring, Coaching	Multiple Probe Multiple Baseline	Increase BSP by teachers' on-task bx increased. After tier the faculty training, results show that the participants' BSP did not increase as intended. Also, one participant did not

Citation	**Teacher Participant**	**Student Participant**	**Setting**	**DV**	**IV**	**Design**	**Results**
							attend scheduled coaching sessions, due to absence.

Table 2

Quality Indicators of Group Experimental and Quasi-Experimental Research Design

	Gregory et al. (2013)	Matsumara et al.(2010)	Reinke et al. (2014)	Ruble et al. (2013)	Podhajski et al. (2009)	Sailors & Price (2015)
Participants	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Sufficient info provided to determine disability/difficulties	N/A	Yes	Yes	Yes	Yes	Yes
Relevant characteristics comparable across conditions	Yes	Yes	Yes	Yes	Yes	Yes
Sufficient info characterizing interventionist/teachers. Comparable across conditions.	Yes	Yes	Yes	Yes	Yes	Yes
Implementation of Intervention and Description of Comparison Conditions	**No**	**No**	**No**	**Yes**	**No**	**Yes**
Intervention clearly described and specific	Yes	Yes	Yes	Yes	Yes	Yes
Fidelity of implementation described and assessed (coaching)	No	No	No	Yes	No	Yes
Comparison condition described	Yes	Yes	Yes	Yes	Yes	Yes
Outcome Measures	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Multiple measures closely aligned with intervention	Yes	Yes	Yes	Yes	Yes	Yes

	Gregory et al. (2013)	Matsumara et al.(2010)	Reinke et al. (2014)	Ruble et al. (2013)	Podhajski et al. (2009)	Sailors & Price (2015)
Outcomes for capturing the intervention's effect measured at appropriate times	Yes	Yes	Yes	Yes	Yes	Yes
Data Analysis	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Analysis techniques linked to research questions	Yes	Yes	Yes	Yes	Yes	Yes
Did the research report include inferential statistics and effect size calculations	Yes	Yes	Yes	Yes	Yes	Yes

Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children*, *71*, 149-164.

Table 3

Quality Indicators of Single Subject Design

	Bethune & Wood (2013)	Coddi ng, et al. (2005)	Duchaine et al. (2011)	Kretlow et al. (2012)	Kretlow et al. (2011)	Martens et al. (1997)	Miller et al. (1991)	Simonsen et al. (2010)	Sutherland et al. (2000)	Thompson et al. (2012)
Participants	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Replicable	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sufficient description of participants	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

	Bethune & Wood (2013)	Coddi ng, et al. (2005)	Duchaine et al. (2011)	Kretlow et al. (2012)	Kretlow et al. (2011)	Martens et al. (1997)	Miller et al. (1991)	Simonsen et al. (2010)	Sutherland et al. (2000)	Thompson et al. (2012)
Sufficient description of setting	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dependent Variable	**Yes**	**No**	**Yes**	**Yes**	**Yes**	**No**	**No**	**Yes**	**Yes**	**Yes**
DV replicable	Yes	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Quantifiable	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Measurement valid and replicable	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Measurement repeated	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
IOA reported	Yes	No	Yes	Yes	Yes	No	No	No	Yes	Yes
Independent Variable	**Yes**	**No**	**Yes**	**Yes**	**Yes**	**No**	**No**	**No**	**Yes**	**Yes**
IV replicable	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Systematically manipulated	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Procedural fidelity measured and described	Yes	No	Yes	Yes	Yes	No	No	No	Yes	Yes

	Bethune & Wood (2013)	Coddi ng, et al. (2005)	Duchaine et al. (2011)	Kretlow et al. (2012)	Kretlow et al. (2011)	Martens et al. (1997)	Miller et al. (1991)	Simonsen et al. (2010)	Sutherland et al. (2000)	Thompson et al. (2012)
Baseline	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**No**	**No**	**Yes**	**Yes**	**No**
Conditions replicable	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Baseline pattern prior to intervention	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes	No
Experimental control/intern al validity	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**No**	**No**	**Yes**	**No**	**Yes**
3 demonstrations of effect	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	No	Yes
Design controlled threats to internal validity	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
Visual analysis	Yes	Yes	Yes	Yes	Yes	No	No	Yes	No	Yes
External Validity	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**No**	**Yes**
Replicated across participants	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes

	Bethune & Wood (2013)	Coddi ng, et al. (2005)	Duchaine et al. (2011)	Kretlow et al. (2012)	Kretlow et al. (2011)	Martens et al. (1997)	Miller et al. (1991)	Simonsen et al. (2010)	Sutherland et al. (2000)	Thompson et al. (2012)
Social validity	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**No**	**Yes**	**No**	**Yes**
DV socially important	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes
Change in DV due to intervention	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes
IV cost effective	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
IV implemented over time	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children*, *71*, 165-179.

Table 4
Coaching Components

	Modeling	Prior training	Pre-observation meeting	E-mail Feedback	Hand written Feedback	Face to Face Feedback	Teacher Goal Setting	Location of coaching
Bethune & Wood (2013)	Yes	Yes	Yes	No	No	Yes	No	Classroom
Codding et al. (2005)	Yes	Yes	No	No	Yes	Yes	No	Outside Classroom
Duchaine et al. (2011)	Yes	Yes	Yes	No	Yes	Yes	Yes	Not Mentioned
Gregory et al. (2013)	No	Yes	No	No	Yes	No	No	Web/Telephone
Kretlow et al. (2012)	Yes	Yes	Yes	No	No	Yes	No	Not Mentioned
Kretlow et al. (2011)	Yes	Yes	Yes	No	No	Yes	No	Not Mentioned
Martens et al. (1997)	No	No	Yes	No	Yes	Yes	Yes	Not Mentioned
Matsumara et al.(2010)	Yes	Yes	No	No	No	Yes	No	Not Mentioned
Miller et al. (1991)	No	No	No	No	Yes	Yes	Yes	Not Mentioned
Podhajski et al. (2009)	Yes	Yes	No	No	No	Yes	No	Not Mentioned
Reinke et al.(2014)	No	Yes	No	No	No	Yes	No	Not Mentioned
Ruble et al. (2013)	No	No	No	No	Yes	Yes	No	Web
Sailors & Price (2015)	Yes	Yes	Yes	Yes	No	Yes	No	In Classroom
Simonsen et al. (2010)	No	Yes	Yes	Yes	Yes	Yes	Yes	Not Mentioned
Sutherland et al (2000)	No	Yes	Yes	No	No	Yes	No	Not Mentioned
Thompson et al. (2012)	No	Yes	No	No	No	Yes	No	Not Mentioned