Improving Instructional Practice through Peer Observation and Feedback

Brady L. Ridge
Alyson L. Lavigne
Utah State University
United States

Citation: Ridge, B. L., & Lavigne, A. L. (2020). Improving instructional practice through peer observation and feedback. Education Policy Analysis Archives, 28(61).
https://doi.org/10.14507/epaa.28.5023 This article is part of the special issue, Policies and Practices of Promise in Teacher Evaluation, guest edited by Audrey Amrein-Beardsley.

Abstract: The Every Student Succeeds Act provides an opportunity for policymakers and researchers to revisit what is known about effective teacher evaluation practices to make better-informed decisions moving forward. Principals—responsible for implementing new teacher evaluation reforms and accommodating the demands to spend more time observing and providing feedback to teachers—are overworked. They have little time to provide high-quality feedback, and may lack important content-based expertise. With these considerations in mind, we explore the role of peer observation and feedback as a vehicle to move beyond high-stakes evaluation and re-center efforts on instructional improvement. Our systematic review of extant literature ($n = 38$ documents, 92% peer-reviewed empirical articles) indicates that peer observation and feedback is a promising
practice for instructional improvement, but one that lacks sufficient evidence. Policy, thus, can encourage innovation and research around this practice so that peer observation and feedback models can be piloted and the most effective established, as well as strategies to tackle the biggest barriers schools, particularly U.S. schools face in implementing such a practice—time.

**Keywords:** coaching; performance; peer coaching; instructional improvement; professional development

**Mejora de la práctica educativa a través de la observación y retroalimentación de pares**

**Resumen:** La Every Student Succeeds Act brinda una oportunidad para que los formuladores de políticas e investigadores revisen lo que se conoce sobre prácticas efectivas de evaluación docente para tomar decisiones mejor informadas en el futuro. Los directores, responsables de implementar nuevas reformas de evaluación docente y de satisfacer las demandas de pasar más tiempo observando y brindando retroalimentación a los docentes, están sobrecargados. Tienen poco tiempo para proporcionar comentarios de alta calidad y pueden carecer de experiencia importante basada en contenido. Con estas consideraciones en mente, exploramos el papel de la observación y retroalimentación entre pares como un vehículo para ir más allá de la evaluación de alto riesgo y volver a centrar los esfuerzos en la mejora de la enseñanza. Nuestra revisión sistemática de la literatura existente (n = 38 documentos, 92% de artículos empíricos revisados por pares) indica que la observación y retroalimentación de pares es una práctica prometedora para la mejora de la instrucción, pero que carece de evidencia suficiente. La política, por lo tanto, puede alentar la innovación y la investigación en torno a esta práctica para que los modelos de evaluación y retroalimentación de pares puedan ponerse a prueba y establecerse de la manera más efectiva, así como estrategias para abordar las barreras más grandes que enfrentan las escuelas, particularmente las escuelas de EE. UU.

**Palabras clave:** coaching; actuación; capacitación; mejora educativa; desarrollo profesional

**Melhorar a prática instrucional através da observação e feedback de colegas**

**Resumo:** A Every Student Succeeds Act oferece aos formuladores de políticas e pesquisadores uma oportunidade de revisitar o que se sabe sobre práticas eficazes de avaliação de professores para tomar decisões mais bem informadas no futuro. Os diretores - responsáveis por implementar novas reformas de avaliação de professores e acomodar as demandas para passar mais tempo observando e fornecendo feedback aos professores - estão sobrecarregados. Eles têm pouco tempo para fornecer feedback de alta qualidade e podem não ter conhecimentos importantes baseados em conteúdo. Com essas considerações em mente, exploramos o papel da observação e feedback de colegas como um veículo para ir além da avaliação de alto risco e re-centrar os esforços na melhoria instrucional. Nossa revisão sistemática da literatura existente (n = 38 documentos, 92% artigos empíricos revisados por pares) indica que a observação e o feedback dos pares são uma prática promissora para o aprimoramento instrucional, mas que carece de evidências suficientes. A política, portanto, pode incentivar a inovação e a pesquisa em torno dessa prática, para que os modelos de observação e feedback por pares possam ser pilotados e o mais eficaz seja estabelecido, bem como estratégias para enfrentar as maiores barreiras que as escolas, principalmente as americanas, enfrentam na implementação de tal prática - tempo.

**Palavras-chave:** coaching; desempenho; treinamento em grupo; melhoria instrucional; desenvolvimento profissional
Improving Instructional Practice through Peer Observation and Feedback

We have known for some time that teachers matter and that teachers are the most important in-school factor that impacts student learning (Aaronson, Barrow & Sanders, 2007; Brophy & Good, 1986; Goldhaber & Brewer, 1999; Good, Biddle, & Brophy, 1975; Konstantopoulos, 2014; Rubie-Davies, 2014). Therefore, it is vital for all students to be taught by effective teachers (Lavigne & Good, 2012). However, how to measure the effectiveness of teachers continues to evolve. Lavigne and Good (2019) note that many schools have turned to high-stakes testing as a means to measure teacher effectiveness, evaluate teachers, and make personnel decisions. Another way that schools attempt to ensure effective teachers for all students is by improving instruction through observation and feedback, practices characterized as supervision. One source of feedback, and the most utilized, particularly in the United States, is administrator-to-teacher feedback, while another practice is peer-to-peer feedback. Given the challenges of administrator-to-teacher feedback, which we document in detail below, in this paper we conduct a systematic review of the literature to examine the utility of peer feedback as a suitable practice to assist instructional leaders in improving instruction. To start, we provide a brief history of teacher evaluation policy. Then, we summarize the research related to commonly used and recent teacher evaluation practices. These two sections provide the rationale for the current review.

Teacher Evaluation: Policy and Research

Teacher evaluation reform has spanned the globe (see the special issue, Global Perspectives on High-Stakes Accountability Policies in Education Policy Analysis Archives guest edited by Holloway, Sørensen, & Verger, 2017). In the United States, 2009 was a particularly transformational year in teacher evaluation reform as Race to the Top (RTTT) was launched. This competitive program allocated more than $4 billion to states to improve instruction, in part, through more effective teacher evaluation. Although only 18 states and the District of Columbia were awarded RTTT dollars, 45 submitted applications. As a result, the ripple effect of RTTT reached teacher evaluation models in nearly all states (Howell, 2015; National Council on Teacher Quality [NCTQ], 2017). These new teacher evaluation models placed a greater emphasis on student achievement growth (with an uptick in the use of value-added models), required that principals spend more time in classrooms observing and providing feedback to teachers, and often included an expanded rating scale as opposed to a dichotomous scale (e.g., effective, ineffective). In some cases, high-stakes were attached to teachers’ evaluation ratings (e.g., hiring, firing, tenure).

Now a decade post-RTTT, some research has pointed to the value of teacher evaluation. These benefits include: increases in student achievement as a function of replacing teachers, particularly replacing low-performing teachers (Adnot, Dee, Katz, & Wyckoff, 2017), as well as for those teachers who have remained (Dee & Wyckoff, 2015). Notably, these findings were

---

Notably, Adnot et al. (2017) conducted a quasi-experimental design in the District of Columbia Public Schools (DCPS). Increases in student achievement as a function of teacher replacements was significant in mathematics (.08 of a standard deviation), but not in reading. Effects were significant and larger for teachers who were replaced due to low performance (i.e., 0.14 SD in reading and 0.21 SD in math). The teacher evaluation system in DCPS was significantly high-stakes (both in terms of rewards and dismissals linked to teacher evaluation scores) and data were not available to determine whether or not these trends were different than those prior to the implementation of IMPACT. It is possible, furthermore, that IMPACT had relatively little impact on the retention of high-performing teachers as the attrition rates of these teachers in DCPS mirrored those rates observed in other urban districts.
documented in one of the most high-stakes teacher evaluation systems implemented post RTTT—the District of Columbia Public Schools—and may not generalize to other teacher evaluation systems.

Meanwhile, other research findings indicate that a greater emphasis on student achievement and an increased number of required observations has not improved teaching and learning in the United States (Stecher et al., 2018). Perhaps this is because even under the best conditions when principals were prepared and had the skills to do teacher evaluation well, they lacked time (Donaldson & Woulfin, 2018; Goldring et al., 2015; Kraft & Gilmour, 2016; Lavigne & Chamberlain, 2017; Stecher et al., 2018). Principals coped by completing fewer observations than designated by policy or cutting observations short, and were unavailable to address teacher concerns (Donaldson & Woulfin, 2018; Stecher et al., 2018). Principals spent more time writing their evaluations than observing teachers and providing teachers with rich feedback (Flores & Derrington, 2017). Evaluating teachers outside of their own content expertise or having limited teaching experience meant that some principals struggled to provide teachers with content-based and specific feedback (Kraft & Gilmour, 2016). Subsequently, only half of teachers indicated that the feedback they received from their principals was useful (Cherasaro, Brodersen, Reale, & Yanoski, 2016). Even the increased number of observations, if they were accomplished by principals, was not enough to reliably measure teacher effectiveness for informing personnel decisions (Hill, Charalambous, & Kraft, 2012; Ho & Kane, 2013; Van der Lans, Van de Grift, Van Veen, & Fokkens-Bruinsma, 2016). Likewise, the increased emphasis on student achievement, particularly the use of value-added measures, has proven to be highly flawed in accurately capturing a teacher’s “true” effectiveness, with high error rates in classifying teachers, even for teachers with 10 years’ worth of data (see Baker et al., 2010, for a comprehensive overview of the concerns in using student achievement data to evaluate teachers). In short, some have concluded that teacher evaluation reform efforts under RTTT have failed to improve teaching and learning (Lavigne & Good, 2019).

On the coattails of RTTT, the Every Student Succeeds Act (ESSA) provides an opportunity to redesign teacher evaluation systems to consider other possibilities. In particular, the ESSA, passed in 2015 reduces the requirement of student growth and no longer requires states to have a teacher evaluation model. This has resulted in states pulling back on RTTT-inspired teacher evaluation by diminishing the weight of growth in teacher evaluation systems, eliminating it entirely, and/or allowing districts, rather than the state, to determine their teacher evaluation systems (Croft, Guffy, & Vitale, 2018). This opportunity, though, raises the question: What did we learn from RTTT teacher evaluation models? What do we know now about the most effective practices in teacher evaluation?

Much of the last decade of research on teacher evaluation has examined high- versus low-stakes policies and practices (Holloway et al., 2017), with various scholars raising concerns about the use of high stakes in teacher evaluation (see, for example, special issue in Teachers College Record edited by Lavigne, Good, & Marx, 2014). While this comparison has real and important implications for schools, principals, teachers, and the students they serve, Holloway et al. (2017) suggest that this dominant debate may restrict the extent to which teacher evaluation research can advance practice and policy.

With that in mind, we examine a re-occurring issue inherent in both low- and high-stakes models—the tension between the dual purposes of teacher evaluation and teacher supervision (see Hazi & Rucinski, 2009 for a review). In the United States, in most schools the building principal will conduct formative observations throughout the year to provide the teacher with non-evaluative feedback.

---

2 Principals’ lack of time has been well documented across multiple decades (see, for example, Kersten & Israel, 2005; OECD, 2019).
feedback for improving practice (supervision) before conducting a summative, end-of-year evaluation. Importantly, as Hazi and Rucinski note, these two activities—supervision and evaluation—are different. Supervision has the goal of helping teachers develop, whereas evaluation serves a personnel function. Yet, in practice these activities are often synonymous in part due to the fundamental conflict in current teacher evaluation models practiced in the United States where the coach is also the judge. Due to supervision functioning ‘incognito’ or under the guise of teacher evaluation, some have suggested that teacher evaluation be part of the discourse on supervision and vice versa (Hazi, 1994).

These calls align with other revelations related to supervision that have emerged just prior to and under recent teacher evaluation models (see Glanz and Hazi, 2019 and Allen and LeBlanc, 2005 for illustrations that demonstrate many of these issues are not new). For example, despite evidence that instructional leadership, broadly defined, appears to be related to staff perceptions of the school’s environment as well as teacher satisfaction (Horn, Klasik, & Loeb, 2010), there appears to be no positive relationship between principals’ time spent observing and providing feedback to teachers and student learning outcomes (Grissom, Loeb, & Master, 2013; Horng et al., 2010). This may be, in part, because the common practice of using a single individual—the principal—to conduct most, if not all observations (Cherasaro et al., 2016) does not align with findings which recommend three observations by multiple individuals to acquire adequate reliability for providing feedback (.70; Hill et al., 2012; Ho & Kane, 2013) and 10 observations for adequate reliability for promoting or dismissing a teacher (.90; Van der Lans et al., 2018), as noted by Lavigne and Good (2019). In the current structure of American schools and the demands placed on the primary evaluator—the principal—ten observations by multiple observers is not feasible.

**Peer Observation and Feedback: A Promising Practice?**

However, providing teachers with reliable feedback from three different observers on three different occasions could be possible through peer observation and feedback. This is a practice districts can leverage and that is used across the globe, but that is underutilized in the United States. Notably, data from the Teaching and Learning International Survey (TALIS) indicates that teachers in other TALIS-participating countries are more likely to receive feedback from peers (42%) than teachers in the United States (27%; OECD 2014a, 2014b). Furthermore, in Ford, Urick, and Wilson’s (2018) examination of the TALIS 2013 data, teacher satisfaction was generally higher when the primary evaluator was a fellow teacher, mentor, or other member of school management (not the principal). Perhaps teachers perceive feedback to be less threatening when it is delivered by a peer (Joyce & Showers, 1982). Together these findings raise the possibility of engaging fellow teachers in formative and/or summative aspects of teacher evaluation.

There have been two notable reviews conducted on the literature on peer coaching (Ackland, 1991; Lu, 2010), however, the review conducted by Ackland was not systematic and

---

3 It is not unusual for a principal to have an evaluation load of 20 teachers/year, which would equate to coordinating nearly 200 observations (10 observations x 20 teachers) by 10 different observers in a single school year.

4 Feedback from teachers (as reported by teachers in lower secondary schools) was more frequently reported by teachers in Korea (84%), Denmark (58%), Latvia (58%), the Netherlands (57%), and Norway (57%).

5 This review was based on the current literature at the time (sources were published from 1983 to 1989). The authors identified 11 sources on expert coaching (which would not be included in this review based on our exclusion criteria) and 18 on peer coaching, but it did not seek to gather any consensus to the utility or effectiveness of the practice, nor was inclusion or exclusion criteria described to help authors understand the scope of the studies included in the review.
written primarily to provide a descriptive account for practitioners about what peer coaching entails, various models for peer coaching (e.g., expert, reciprocal), and how to implement peer coaching in schools. Lu did conduct a systematic review of studies on peer coaching from 1997 to 2007 ($N = 8$ studies), however, it was focused on peer coaching in pre-service contexts. Results from the review indicate that peer coaching is a promising practice for pre-service teacher growth and development when student teachers receive training on peer coaching. However, the author did not extend these conclusions to consider how pre-service peer coaching might be applied (or not) or support the extension of peer coaching to practicing teachers.

With these considerations in mind, we review extant literature on peer observation and feedback. We do this with the underlying assumption that peer observation and feedback may be a useful vehicle to move beyond the high- versus low-stakes debate and instead to center instructional improvement, and emphasize supervision within teacher evaluation models. To address the limitations of and extend upon the findings from prior reviews (Ackland, 1991; Lu, 2010) we conduct a systematic review of the literature on pre-service and in-service peer coaching so that we might consider how these two bodies of literature might inform one another.

**Methodology**

**Defining Peer Observation and Feedback**

Peer observation and feedback is often subsumed under the larger umbrella of peer coaching. Robbins (2015) defines peer coaching as:

- a powerful, confidential, non-evaluative process through which two or more colleagues work together to: reflect upon and analyze teaching practices and their consequences; develop and articulate curriculum, create informal assessments to measure student learning; implement new instructional strategies, including the integrated use of technology; plan lessons collaboratively; discuss student assessment data and plan for future learning experiences; expand, refine, and build new skills; share ideas and resources; teach one another; conduct classroom research; solve classroom problems or address workplace challenges; and examine and study student learning with the goal of improving professional practice to maximize student success. (p. 9)

Whereas instructional coaches often exit their own classroom to oversee other classroom teachers, peer coaches typically hold the same ‘rank and status’ and are heavily focused on collaboration. Thus peer coaches or observers have not traditionally assisted in teacher evaluations, but rather provide formative feedback throughout the school year. Teachers that have peer-to-peer coaching models in place can increase teacher collaboration, can increase the observations of one another, as well as receive and provide feedback teachers receive in order to improve instructional practice.

Robbins (2015) organizes peer coaching into two categories: collaborative work and formal coaching. In the former—collaborative work—professional colleagues use collaborative structures to promote learning, generally, but not in relationship to specific observations of classroom practice. In the latter—formal coaching—classroom observations are key, including pre- and post-observation conferences. This type of peer coaching typically centers around a specific lesson and the learning outcomes it produced.

These activities often fall under the larger umbrella of supervision—what Glickman, Gordon, and Ross-Gordon (2017) broadly define as “assistance for the enhancement of teaching and learning” (p. 9). In their view of collegial supervision, supervision is: not a hierarchical
relationship between teachers and supervisors; the province of teachers and formally designated supervisors; focused on teacher growth instead of compliance; a means of facilitating teacher collaboration; and grounded in ongoing reflective inquiry (p. 7). Supervision, then, can include experienced teachers who also serve as mentors, clinical supervision programs (like the peer coaching discussed above), teacher leaders who receive release time to observe and support other teachers, as well as collegial peer-coaching pairs and triads (also referred to later as reciprocal peer coaching).

**Search Criteria**

We used the above definitions to conduct preliminary searches. Based on our preliminary searches, we selected the following search terms to identify appropriate literature (peer-reviewed, empirical journal articles, dissertations, and research reports published in English) for inclusion: professional development AND (“peer coaching” or “instructional coaching”).

This search yielded 676 results from the combined search using the following databases: EBSCOhost, ERIC, Education Source, and the Professional Development Collection respectfully. Notably, no time period was defined for the review of literature, but the initial search yielded results dating as far back as 1971.

**Inclusion and exclusion criteria.** In Phase 1, duplicates were removed ($n = 135$). Then, we reviewed the article abstracts for relevance. We carefully reviewed all of the abstracts of the remaining 541 documents applying the definition of peer-to-peer observation and feedback, described above, to guide exclusion and inclusion decisions. Phase 1 resulted in the exclusions of 511 documents for the following reasons: lack of relevance, instructional coaching or literacy coaching (where the individual holds a pseudo administrative role) as opposed to peer-to-peer observation and feedback or more formal peer coaching arrangements.

In Phase 2, we randomly assigned documents to an individual author and each author read their respective articles in depth. In Phase 2, we excluded ten documents for the following reasons: instructional coaching or literacy coaching (where the individual holds a pseudo administrative role) as opposed to peer-to-peer observation and feedback, or the document was not empirical research. Documents which were not empirical were typically practitioner-oriented manuscripts that provided suggestions on how to apply peer-to-peer coaching as opposed to original, empirical studies of peer coaching. Lastly, we removed literature reviews to avoid falsely giving more weight to a certain finding as it would be possible that studies cited in any given literature review may re-appear again in our own review. At the conclusion of Phase 2, 20 documents remained.

Greenhalgh and Peacock (2005) argue that in order to provide a robust illustration of a body of literature on a particular topic, systematic reviews cannot rely solely on the results acquired from predefined search protocols. Thus, in Phase 3, we applied the backward snowballing method, to identify high quality sources that would not otherwise be identified in using predefined search protocols. We reviewed the introductions, literature reviews, and methodologies for each of the 20 documents identified in Phase 2 to identify any possible relevant literature. Duplicates were eliminated, and then the titles of these references were examined. If still deemed relevant, we

---

6 Despite earlier rationale for possibly excluding instructional coaching, initial searches of peer coaching revealed such a limited number of documents, the search terms were expanded to include instructional coaching to account for instances where instructional coaching may have been used to describe documents that otherwise encompass peer coaching processes.

7 This does not necessarily signify that peer coaching did not exist before 1971, but that this terminology using our search parameters did not appear in the literature until 1971.
acquired their abstracts. We collected abstracts, when available, for 63 sources. We then followed the review steps described in Phase 1, which resulted in 18 sources being added to the final pool. Our final sample, thus, consisted of 38 documents which met the criteria for inclusion. The included literature spans a period of 35 years from 1984 to 2019 (although we did not set any exclusion dates); 37% of the publications were published in the last decade. See Table 1.

**Types of Data**

Informed by notable and exemplary reviews of literature (Snodgrass-Rangel, 2017) as well as Hallinger’s (2014) systematic review of 38 reviews of research in educational leadership, for each article, we collected information on the following questions, questions that guide the organization of our results:

1. What conceptual and/or theoretical framework guides the study?
2. What is the geographic locale?
3. What is the level and context of the study?
4. What is study sample and size?
5. How do the author(s) define peer coaching?
6. What are the data measures or sources?
7. What is the study design?
8. What are the major findings?

These questions were chosen to assist our efforts in providing a rich narrative of extant research on peer observation and feedback. We primarily chose descriptive characteristics to identify common or saturated aspects of the research as well as any gaps in the literature. This descriptive approach also helped informed conclusions about the generalizability of peer observation and feedback beyond the context of the studies by reflecting some of the standards of reporting in the field (American Educational Research Association [AERA], 2006; American Psychological Association [APA], 2020). Finally, although we recognize that research can be guided by key issues, debates, barriers, gaps in the literature, or practical concerns (APA, 2020) we were particularly attentive to the presence or absence of theoretical and conceptual frameworks because of critiques and findings that education research is relatively atheoretical (e.g., Ford, Lavigne, Fiegener, & Si, 2020; Teddlie & Reynolds, 2000; Trujillo, 2013, 2016) and is at risk of becoming even more so, diminishing our ability to reach conclusions about larger patterns in human development and learning (Dimitriadis, 2009).
<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Document Type*</th>
<th>Conceptual/Theoretical Framework/Theorization(s)</th>
<th>Geographic Locale</th>
<th>Level</th>
<th>Context</th>
<th>Sample &amp; Size</th>
<th>Definition of Peer Coaching</th>
<th>Measures</th>
<th>Study Design</th>
<th>Finding(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anderson, Barkdale, &amp; Him (2005)</td>
<td>J</td>
<td>None mentioned</td>
<td>United States</td>
<td>Primary</td>
<td>34 elementary pre-service teachers</td>
<td>No clear definitions; implicit in a peer-to-peer process</td>
<td>Weekly Journals, lesson plans, open-ended surveys, and observations</td>
<td>Qualitative</td>
<td>Students enjoyed their experience observing and being observed.</td>
<td></td>
</tr>
<tr>
<td>Arnou, Kahre &amp; Krugkamp (2004)</td>
<td>J</td>
<td>None mentioned</td>
<td>United States</td>
<td>Secondary</td>
<td>14 veteran teachers</td>
<td>“Peer coaching is a practice in which teachers work together through a process that includes pre-conferencing, observation, and post-conferencing.” (p. 26)</td>
<td>Interviews</td>
<td>Qualitative</td>
<td>Teachers reported the following reasons for participating in peer coaching: desire to learn, informal peer coaching experience, need for meaningful feedback, attraction to choice, dissatisfaction with traditional observation, meaningful feedback, self-directed learning, trust among peers, worth from extra work; increased morale in peer coaches.</td>
<td></td>
</tr>
<tr>
<td>Ben-Peretz, Gottlich &amp; Guldanov (2010)</td>
<td>J</td>
<td>Israel</td>
<td>2 in-service teachers</td>
<td>of continuous improvement</td>
<td>Qualitative</td>
<td>Interviews</td>
<td>Teacher has positive perceptions of what collaboration can do to improve teaching.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Betem &amp; Anderson (2010)</td>
<td>J</td>
<td>Development/Effective Model of Teaching</td>
<td>United States</td>
<td>Secondary</td>
<td>4 pre-service teachers</td>
<td>No clear definition; implicit in a peer-to-peer process</td>
<td>Open-ended interviews</td>
<td>Research</td>
<td>Peer coaching was found to alter practice as reported by those who engaged in peer coaching. Their findings suggest the importance of engaging pre-service teachers early in their program in peer coaching.</td>
<td></td>
</tr>
<tr>
<td>Bruce &amp; Ross (2008)</td>
<td>J</td>
<td>Bandura</td>
<td>Canada</td>
<td>Primary</td>
<td>12 in-service teachers</td>
<td>Observations; Pre- and post assessments</td>
<td>Mixed Method: coaching caused participants to reflect more explicitly.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Castleda-Londoño (2017)</td>
<td>J</td>
<td>None mentioned</td>
<td>Colombia</td>
<td>School</td>
<td>3 in-service teachers</td>
<td>No clear definition; implicit in a peer-to-peer process</td>
<td>Interviews</td>
<td>Qualitative</td>
<td>Observation/feedback as means to understand one’s own self as teacher.</td>
<td></td>
</tr>
<tr>
<td>Author(s)</td>
<td>Year</td>
<td>Country</td>
<td>School Level</td>
<td>Type of Teacher</td>
<td>Methodology</td>
<td>Research Design</td>
<td>Findings</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>------</td>
<td>---------</td>
<td>--------------</td>
<td>----------------</td>
<td>-------------</td>
<td>----------------</td>
<td>----------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jao (2013)</td>
<td>J</td>
<td>The Ten Dimensions of Mathematics Education</td>
<td>Canada</td>
<td>Primary</td>
<td>4 in-service teachers</td>
<td>No explicit definition; implicit definition is from one another (Loudes-Horsley et al., 2003)</td>
<td>Qualitative</td>
<td>The participants found benefits of peer coaching, but also found some challenges. Collaboration was a good aspect of peer coaching. The greatest challenge was the amount of time to peer coach.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Koch (2014)</td>
<td>D</td>
<td>Theory of Social Learning Organization</td>
<td>United States</td>
<td>Secondary Private School</td>
<td>17 in-service teachers</td>
<td>No explicit definition; implicit definition is from one another (Loudes-Horsley et al., 2003)</td>
<td>Quantitative Regression</td>
<td>Teachers who participated in peer coaching experienced more collaboration, collegiality, and teaching self-efficacy.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kohler, Griffler, Shearer, &amp; Good (1997)</td>
<td>J</td>
<td>Learning Organization</td>
<td>United States</td>
<td>Primary</td>
<td>4 elementary teachers</td>
<td>No clear definition; implicit is a peer-to-peer</td>
<td>Single case design</td>
<td>Their results indicated that areas that were not discussed during the seven coaching sessions were rarely implemented or refined in teachers' subsequent instructional practice.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kohler, Ecell, &amp; Paluselli (1999)</td>
<td>J</td>
<td>Learning Organization</td>
<td>United States</td>
<td>Primary</td>
<td>2 Kindergarten teachers</td>
<td>No explicit definition; implicit definition is from one another (Loudes-Horsley et al., 2003)</td>
<td>Interviews Mixed Method</td>
<td>Coaching increased teacher use of suggestions, prompts, and questions in student peer activities and adaptation of materials-on-the-spot. The changes lasted after coaching was concluded (as measured in the maintenance phase).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lee &amp; Choi (2013)</td>
<td>J</td>
<td>Learning Organization</td>
<td>Korea</td>
<td></td>
<td>10 pre-service secondary physical education teachers</td>
<td>No clear definition; implicit is a peer-to-peer</td>
<td>Journals, evaluation forms, interviews</td>
<td>Qualitative</td>
<td>Findings indicated that peer coaching facilitates reflection by helping teachers face barriers and connecting planning and instruction with reflection. Through reflections, peer coaches were able to reframe and accept their practices, confront and extend, and adjust and re-aim their instructional goals.</td>
<td></td>
</tr>
<tr>
<td>Licklider (1995)</td>
<td>J</td>
<td>Learning Organization</td>
<td>United States</td>
<td>Secondary</td>
<td>11 volunteer secondary teachers</td>
<td>No explicit definition; implicit definition is from one another (Loudes-Horsley et al., 2003)</td>
<td>Audio tapes for coaching sessions</td>
<td>Qualitative</td>
<td>Teachers improved on effective questioning techniques; teachers found peer coaching was most helpful when it allowed teachers to practice skills in the classroom, followed by observing a colleague and providing feedback, followed by being observed and receiving feedback.</td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>J</td>
<td>Country</td>
<td>Sample Size</td>
<td>Method</td>
<td>Data Collection</td>
<td>Findings</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>---</td>
<td>---------</td>
<td>-------------</td>
<td>--------</td>
<td>----------------</td>
<td>---------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ma, Xin, &amp; Du (2018)</td>
<td>J</td>
<td>China</td>
<td>None mentioned. 20 in-service teachers (Treatment = 10; Control = 10)</td>
<td>Quantitative-Experimental</td>
<td>Learning participation and pedagogy was higher in treatment group, but design skills were equal.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meng, Tajroemsuk, &amp; Sophas (2013)</td>
<td>J</td>
<td>China</td>
<td>None mentioned. Vygotsky; specifically zone of proximal development</td>
<td>Mixed Method</td>
<td>Significant gains in knowledge from pre to post; teachers found peer coaching to be helpful.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Murray, Ma, &amp; Mazur (2009)</td>
<td>J</td>
<td>United States</td>
<td>None mentioned. Control group had 5 teachers and 105 students; Experimental group had 6 teachers and 202 students. Mathematics</td>
<td>Quasi Experimental (Used surveys and post observation conferences)</td>
<td>Peer coaching was not associated with any improvement in mathematics achievement of students.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neubert &amp; McAllister (1993)</td>
<td>J</td>
<td>United States</td>
<td>None mentioned. 144 pre-service teachers</td>
<td>Qualitative</td>
<td>1. Peer coaching practicum was perceived to be positive - increase theorizing, student thinking about teaching.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Owens (2004) | J | New Zealand | Not explicit | Research | 12 pre-service teachers physical education provide feedback in post lesson discussions. (p. 47) interviews | 1. Peer coaching practicum was perceived to be positive - increase theorizing, student thinking about teaching.

"generally involves two colleagues engaged in a mutually supportive relationship (Neubert & McAllister, 1993). It is a confidential process through which instructors provide one another with assistance, feedback, and support, and share their expertise, for the purpose of enhancing learning" (p. 293).
Table 1, cont.

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Model</th>
<th>Country</th>
<th>Level</th>
<th>Subjects</th>
<th>Sample Size</th>
<th>Methods</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearce, De la Fuente, Hartweg, &amp; Weinburgh (2019)</td>
<td>Trust Theory</td>
<td>United States</td>
<td>Secondary Biology</td>
<td>29 in-service teachers</td>
<td>&quot;the terms mentoring and coaching are often used interchangeably when an experienced teacher is helping a novice teacher. However, there are subtle differences... coaching is generally more task oriented and focused on performance.&quot; (p. 117)</td>
<td>Focus groups and notes (pre and post); observations of peer coaching</td>
<td>Qualitative: Case Study</td>
</tr>
<tr>
<td>Phillips &amp; Glickman (1991)</td>
<td>Matching Model</td>
<td>Georgia</td>
<td>Primary</td>
<td>22 in-service teachers</td>
<td>Conceptual Systems Theory; Developmental</td>
<td>Phillips and Glickman, p. 21)</td>
<td>Paragraph Completion Method (to measure teachers' conceptual level); audiotaped post-observation conferences; interviews (pre and post); Supervisory Practices for Promoting Instructional Improvement survey; end of program teacher survey Evaluation</td>
</tr>
<tr>
<td>Pollara (2012)</td>
<td>Professional Development, Reverse</td>
<td>United States</td>
<td>Primary</td>
<td>15 in-service teachers</td>
<td>&quot;a process in which two or more teachers work towards a specific and determined purpose in order to improve and validate their practices in the classroom.&quot; (p. 175)</td>
<td>Surveys (pre and post); focus groups; interviews</td>
<td>Action Research</td>
</tr>
<tr>
<td>Porras, Díaz, &amp; Nieves (2018)</td>
<td>Peer Coaching</td>
<td>Colombia</td>
<td>Primary</td>
<td>10 in-service elementary teachers</td>
<td>None mentioned</td>
<td>Survey; English proficiency test; field notes; interviews</td>
<td>Mixed Method</td>
</tr>
<tr>
<td>Prince, Snowden, &amp; Matthews (2010)</td>
<td>None mentioned</td>
<td>United Kingdom</td>
<td>Secondary Science</td>
<td>38 pre-service teachers</td>
<td>Peer Coaching</td>
<td>Questionnaires to provide quantitative data</td>
<td>Mixed Method</td>
</tr>
<tr>
<td>Scheidel, Congdon &amp; Stambaugh (2010)</td>
<td>None mentioned</td>
<td>United States</td>
<td>General ed</td>
<td>3 pre-service teachers</td>
<td>Peer Coaching</td>
<td>No clear definition; implicit is a peer-to-peer process</td>
<td>Mixed Method</td>
</tr>
</tbody>
</table>
Table 1, cont.

<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Setting</th>
<th>Sample Size</th>
<th>Findings and Implications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Showers (1984) R</td>
<td>United States</td>
<td>21 teachers and 6 peer coaching in 2 districts</td>
<td>No clear definition; implicit is peer-to-peer</td>
<td>Student data, lesson plans, observations, and interviews; Qualitative. Peer coaching can be trained relatively quickly. Transfer of training was better for coached vs. uncoached teachers. Students of coached teachers scored higher on concept attainment.</td>
</tr>
<tr>
<td>Syh-Jong &amp; Hsiao-Chuan (2009)</td>
<td>Taiwan</td>
<td>Secondary</td>
<td>123 secondary students</td>
<td>No clear definition; implicit is peer-to-peer</td>
</tr>
<tr>
<td>Shiu-Fong &amp; Wing-Shuen (2008)</td>
<td>Hong Kong</td>
<td></td>
<td>35.5 in-service teachers across 2 studies</td>
<td></td>
</tr>
<tr>
<td>Spaulds (1986) J</td>
<td>United States</td>
<td>Secondary</td>
<td>19 junior or high teachers</td>
<td>No clear definition; implicit is peer-to-peer</td>
</tr>
</tbody>
</table>
Table 1, cont.

<table>
<thead>
<tr>
<th>Vacillotto &amp; Cummings (2007)</th>
<th>None mentioned. Brazil</th>
<th>16 graduate student pre-service teachers</th>
<th>No explicit definition implicit definition is reciprocal coaching</th>
<th>Peer lesson plans, peer observation, reflective journals, pre and post surveys, and post interview</th>
<th>Qualitative</th>
</tr>
</thead>
</table>

- "A training method in which pairs of practitioners observe each other and provide consultative assistance in correctly implementing teaching strategies and proposing alternative solutions to recognized instructional needs" (p. 74)

1. Perceived positive impact of the implementation of an Instructional Strategies Model along with a paired peer placement with peer coaching in early elementary practicum on student teachers' experiences and professional growth. 2. Benefits of peer coaching included: development of support and collegiality, greater degree of instructional strategy implementation, and enhanced opportunities for reflection and self-analysis for both pre-service and peer coach.

| Wynn & Kromrey (1999) | None mentioned. United States Primary | Pre-service teachers, 78 continued in second year follow-up | Peer coaching forms; surveys on advantages and disadvantages of peer coaching; weekly dialogue journals; videotapes and interviews | Mixed Method |

- "A training method in which pairs of practicum students observe each other and provide consultative assistance in correctly implementing teaching strategies and proposing alternative solutions to recognized instructional needs." (p. 22)

1. Practicum students' pre-teaching concerns were: discipline, management, and verbal and non-verbal communication. Over time these concerns continued to be focused on survival. 2. 75% of the coaches suggestions focused on survival -- these concerns were equally balanced with curriculum concerns over time (over the course of the semester). 3. Coaches' reflections illustrated a balance between survival and curriculum with pupil/curriculum concerns dominating toward the end of the semester. 4. Post-teaching reflections of practicum students showed the same shift as was the case with lesson goals as well.
Teachers learn when they are intrinsically motivated to take part in professional development programs, when they feel a certain pressure toward experimenting with new instructional methods, and when they are able to discuss their experiences within a safe, constructive, and trustworthy reciprocal peer coaching environment.

* D = dissertation; J = peer-reviewed journal article; R = research report

<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Data Collection</th>
<th>Method</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zwart, Wubbels, Bergen, &amp; Bolhuis (2007)</td>
<td>28 in-service teachers (14 dyads)</td>
<td>No explicit definition implicit</td>
<td>Mixed Method</td>
<td>Teachers learn when they are intrinsically motivated to take part in professional development programs, when they feel a certain pressure toward experimenting with new instructional methods, and when they are able to discuss their experiences within a safe, constructive, and trustworthy reciprocal peer coaching environment.</td>
</tr>
</tbody>
</table>


* D = dissertation; J = peer-reviewed journal article; R = research report
Data Evaluation and Analysis

The authors communicated continuously throughout the article review process to make any necessary updates to the data collection approach and/or procedures, although data were collected and recorded on each article independently. At the conclusion of the initial review, a subset of articles were randomly chosen (n = 4) to be double-coded and to check for reliability. On this randomly chosen subset, authors had exact agreement 100% of the time. Themes highlighted in the findings were reached in consensus after the authors compared their lists of most salient findings from their respective article assignments.

Findings

Conceptual and Theoretical Frameworks

In more than half (63%) of the included literature, there was no mention of a theory or a conceptual framework that guided the study. This could be because the author(s) did not use a framework to guide their study or, it is possible conceptual and theoretical frameworks were used, but not mentioned in the publication. Conceptual and theoretical frameworks and theorists that were noted included: trust theory (e.g., Pearce, De la Fuente, Hartweg, & Weinburgh, 2019), 10 dimensions of mathematics education (e.g., Jao, 2013), Vygotsky (e.g., Bowman & McCormick, 2001; Lee & Choi, 2013; Thijs, & Van den Berg, 2002), conceptual systems theory and developmental matching model (e.g., Phillips & Glickman, 1991), social constructivism, Bandura (e.g., Bruce & Ross, 2008; Licklider, 1995), learning organization theory (e.g., Koch, 2014), the learning community model (e.g., Koch, 2014), and social constructivist theory (e.g., Koch, 2014; Lee & Choi, 2013). Given this variation, there does not appear to be consensus within the included literature on what theoretical framework guides the study of peer-to-peer feedback. However, the social and collaborative nature of the activity of peer-to-peer coaching and feedback made some theoretical frameworks, such as those that purport that knowledge is co-constructed, more appropriate and more frequently cited than others.

Aside from evidence of inconsistent use of theory and conceptual frameworks to inform study design, there was frequent use of seminal works on peer coaching to inform the design or description of peer coaching. The scholarship of Beverly Showers and Bruce Joyce (e.g., Joyce & Showers, 1980, 1981, 1982, 1987, 1988; Showers, 1985; Showers & Joyce, 1996), scholars who have been described as “pioneers of peer coaching” (Anderson, Barksdale, & Hite, 2005, p. 99), was cited in numerous sources included in this review (e.g., Anderson et al., 2005; Bowman & McCormick, 2000; Britton & Anderson, 2010; Hall & McKeen, 1989; Hasbrouck & Christen, 1997; Kohler, Crilley, Shearer, & Good, 2001; Kohler, Ezell, & Paluselli, 1999; Lee & Choi, 2013; Licklider, 1995; Murray, Ma & Mazur, 2009; Neubert & McAllister, 1993; Slater & Simmons, 2001; Sparks, 1986; Syh-Jong & Hsiu-Chuan, 2009; Vacilotto & Cummings, 2007).

Geographic Locale

The studies were conducted in a variety of places across the globe. Of the literature included in this review, geographic locations included: the United States, Korea, Taiwan, China, Canada, the Netherlands, Brazil, Colombia, Botswana, Turkey, Israel, and New Zealand. Even within the most represented geographic region in this review—the United States—and despite the fact that most studies had consistent definitions of peer coaching, the peer coaching models differed widely, making it nearly impossible to derive an understanding of the effectiveness of types of models even across different settings within a single geographic locale.
Level, Context, & Study Sample and Size

Peer-to-peer feedback was not as evident in traditional secondary settings as in other settings such as elementary and early childhood (primary) settings. Most studies at the secondary level were located within a specific school or context (e.g., private school: Phillips & Glickman, 1991; special language school: Castañeda-Londoño, 2017). Furthermore, a number of studies examined content-specific peer coaching (e.g., Mathematics: Jao, 2013; Murray et al., 2009; Science: Thijs & van der Berg, 2002).

Documents included in this review also organized into pre-service and in-service (as it pertained to the study sample). Twelve articles, which comprised almost one-third of the articles in this review of literature (32%), examined peer-to-peer feedback in pre-service teachers. Six studies were conducted in international settings and the remainder in the United States. For example, Wynn and Kromrey (2000) focused on a model where pre-service teachers can help one another during their practicum. In these settings, pre-service teachers typically worked collaboratively with a partner while on practicum to plan lessons, observed each other's teaching, and provided feedback in post lesson discussions (Ovens, 2004, p. 47). Higher education seemed to be better equipped to apply peer-to-peer feedback, in part, because of the autonomy and flexibility such settings provided over traditional K-12 public schools, particularly those located in the United States.

Sample sizes for a majority of studies were small, with the exception of two studies (n = 355 teachers across two studies in Shui-Fong & Wing-Shuen (2008); n = 565 for Hall & McKeen (1991)). In every other case, sample sizes were in the double or single digits. Case studies were not uncommon. For example, Ben-Peretz, Gottlieb, and Gideon (2018) included only two teachers in Israel in their study of peer coaching, while Jao (2013) chose to include four teachers.

Defining and Operationalizing the Independent Variable & Study Design

In a majority of the literature, author(s) explicitly provided a definition of peer coaching. This was particularly important as many of the studies were piloting, implementing, and/or assessing the effectiveness of a peer coaching program. Study designs were primarily qualitative and mixed methods, with only four quantitative studies. Notably, of the four quantitative studies, two were quasi-experimental designs. Given the difficulty of implementing peer coaching programs, particularly in public school settings, it is not surprising that methodologies that are more powerful with smaller sample sizes dominated the literature. Furthermore, even quasi-experimental studies were generally small in regards to sample size. Study designs drove the operationalization of the independent variable and the measures and sources of data used for analysis. As such, interviews, focus groups, observations, audio- and video-recordings, were commonly used, as well as surveys assessing attitudes about peer coaching and feedback. As expected, pre- and post- assessments were common in studies that sought to determine the effect of peer coaching on teachers (e.g., Bruce & Ross, 2008; Pollara, 2012).

Findings: Themes

Collectively, the studies illuminated various benefits of peer coaching, including, but not limited to increased: knowledge (Meng, Tajaroensuk, & Seepho, 2013; Porras, 2008), opportunities to practice and refine instructional skills and goals (Lee & Choi, 2013; Licklider, 1995), classroom management skills (Pollara, 2002), use of common planning time (Pollara, 2002), reflection as measured by frequency (Gonen, 2016) and quality (Bruce & Ross, 2008; Lee & Choi, 2013), and implementation of reform expectations for instruction (Bruce & Ross, 2008). One study documented that peer coaching was an effective approach for changing reported instruction for mathematics teachers (Thijs & van den Berg, 2002), while another study indicated found that peer
coaching was associated with no significant improvement in mathematics achievement (Murray et al., 2009). However, what follows are two aspects that were particularly salient in the research and that have important implications for practice and policy—collaboration and conditions.

**Collaboration.** Collaboration was the most prominent theme from the literature. Various articles specifically mentioned collaboration among teachers as one of the major reported benefits of peer-to-peer feedback (Jao, 2013; Koch, 2014; Phillips & Glickman, 1991; Pollara, 2012; Porras, Diaz, & Nievens, 2018). This was measured and defined in different ways. For example, Hall and McKeen (1989) measured “interaction” or as they defined it, the frequency in which teachers engaged in a variety of activities during their peer coaching. This included the frequency in which teachers reported they, “make collective agreements to test an idea” and “prepare lesson plans with other teachers” (38% and 42% rated the frequency of these activities as a 3 or higher with 5 = frequently). Others examined content and nature of interactions between teachers during peer coaching. For example, Murray et al. (2009) examined what teachers discussed (and how) in post-observation conferences and found that few questions were asked and few compliments were provided during post-observation conferences. Teachers often provided descriptive statements as opposed to analysis, using a positive, supportive tone. Finally, in these post-observation conferences teachers shared the discussion, resulting in relatively equal talk time for both teachers. In alignment with this latter finding, Meng and colleagues (2013) maintain that peer-to-peer feedback is mutually beneficial. It helps the observed as well as the observers. For example, Pollara (2012) asserts that when peer collaboration increases, teacher isolation is reduced, an aspect that has often characterized the teaching profession. Thus, peer coaching creates an environment of teachers working together to solve meaningful problems, which may, in turn, improve teachers’ self-efficacy (Bruce & Ross, 2008).

**Conditions.** In understanding the relatively unique context of each study, we became acutely aware that many of these studies included convenience samples or sites. Furthermore, there emerged antecedent conditions or environments that were particularly ripe for the implementation of a peer coaching program, which in turn, may impact a peer coaching program’s success. For example, in several of the studies, research found buy-in, trust, or willingness to participate in peer-to-peer feedback, to be an important part of the program’s success (Castañeda-Londoño, 2017; Lam, & Lau, 2008; Pollara, 2012). Thus, it might be difficult to pair teachers with people they do not trust. It might also be difficult to change the perception of teachers who do not want to participate in new professional development, although this is a challenge of all professional development and not just peer-to-peer feedback. Furthermore, training emerged as an important antecedent variable as various studies trained participants on peer coaching techniques prior to implementation (e.g., Britton & Anderson, 2009; Neubert & McAllister, 1993). Thus, the quality of the training becomes an important factor when understanding and examining the effectiveness of peer coaching.

**Discussion**

Hypothetically, peer coaching boasts a number of benefits. Licklider (1995) describes these well:

When teachers prepare for a dialogue with a colleague about their own teaching, they must reflect about what they chose to do and why. They must also think about the effectiveness of their choice of behaviors and be ready to discuss the future uses of certain techniques and strategies. When teachers prepare to give feedback to a peer coaching partner, they must reflect about the use of a teaching technique in a different
way than they do when merely observing teaching without the obligation of feedback. They are, for example, forced to think about the appropriateness of the technique in the context in which it was used in the classroom. They have to consider how well it worked and why. They have to think deeply about how to provide the feedback and how to answer questions that a peer coaching partner might raise. (p. 57)

Licklider proposed that this high-level reflection on teaching practices might foster more profound changes in instruction. In this review, we set out to examine the evidence in support of this hypothesis, among others (e.g., actual change in instructional practice) in the extant literature on peer observation and feedback. We did so primarily to determine how this literature might be used to inform policy and practice as it pertains to teacher evaluation. Despite the challenges principals face in observing and providing teachers with high-quality feedback and the positive perceptions teachers hold about peer feedback, overall, our findings from our review of 38 studies indicate that this body of literature does point to various benefits. However, it does not provide adequate evidence to advocate for or against peer observation and feedback, on a global level or even more locally, and specifically as a way to improve instructional practice on a school- or district-wide scale. This may shift with the onset of more empirical studies, particularly quasi experimental observational studies on the effect of peer observation and feedback (as opposed to a supervisor, external observer, or even other approaches to improve instructional practice) on improving actual classroom practice, and subsequently, student achievement.

Limitations

This review is not without limitations. In any review of the literature, search inclusion and exclusion criteria as well as search terms can expand and restrict the final pool of included literature and therefore plays a strong role in the content and validity of the results. For this review, we made a number of intentional decisions in regards to inclusion and exclusion criteria in order to strengthen the validity of our conclusions. First, in our initial exploration of conducting this review, it was acutely apparent research on this topic was limited. This indicated that we could cast a relatively wide net without forfeiting the integrity of the review. In doing so, then, we did not set publication date exclusion or inclusion criteria. Also and informed by OECD (2014) results in the use of peer observation and feedback on a more global scale, we did not set any geographic criteria. Third, we intentionally chose to include research on both pre-service and in-service teachers knowing that pre-service settings and teachers differ in important ways from in-service settings and teachers, but that we may garner a better understanding of the potential of the practice by including studies that may not face the same barriers of time and scheduling that are generally experienced by in-service teachers. Finally, we engaged in backward snowballing (Wohlin, 2014) to account for the limitations of relying solely on a pre-determined and prescribed search criteria (Greenhalgh & Peacock, 2005). We are acutely aware that relevant research is sometimes still missed (e.g., Burgess, Rawal, & Taylor, 2019; Papay, Taylor, Tyler, & Laski, 2016), despite our efforts to engage in a robust and high-quality search.

Despite these efforts, excluding studies not published in English may have resulted in the selection of international studies that are not representative of the extent international research on peer coaching as a whole. Furthermore, although we piloted our search criteria and revised it as needed prior to conducting the full review, it is possible that we failed to include studies that would

For example, after multiple initial search attempts, we realized that including the term “professional development” would increase the probably that we would acquire in our search results the most applicable studies.
otherwise meet our criteria. For example, as it pertains to studies conducted across the globe, we imagine there may exist variations in programs and terminology that may be relevant to our review such as lesson study or peer supervision.

Implications

Even with these abovementioned limitations in mind, our review of the literature on peer coaching and feedback to assess its feasibility as a possible practice of promise demonstrates large gaps in the research, perhaps a large gap between practice and research\(^9\), and an even larger gap as it pertains to research, practice, and policy. We provide extended comment below of the implications of our findings for research, practice, and policy.

**Research.** The literature, as a body of knowledge, is still relatively young\(^10\), underdeveloped, yet, we reiterate—promising. For example, in all of the included literature, except three studies (Murray et al., 2009; Shui-Fong & Wing-Shuen, 2008; Zwart, Wubbels, Bergen, & Bolhuis, 2009\(^11\)) findings indicate that peer coaching is positive. Many studies that document the positive benefits of peer coaching are very small case studies, or observations of an approach implemented in a local school or district. Even in academic journal outlets, the level of reporting often does not meet adequate standards for reporting research results (see, for example, AERA, 2009). Furthermore, despite some quasi-experimental designs as well as some use of pre- and post-measures, relatively few studies measured observed change in *instructional practice* as an outcome or dependent variable (see Bowman & McCormick, 2001; Kohler, Ezell, & Paluselli, 1999; Murray et al., 2009, for more rigorous study designs). The implications of these gaps in the existing research naturally suggest that more research is needed. We provide specific suggestions below that might enhance the existing body of research in meaningful ways.

Ultimately, the goal would be to extend the existing research to determine how peer coaching and feedback is effective—*for whom and under what conditions*. Notably, some research has suggested that the effects of teacher evaluation may be more salient for teachers who may not have experienced evaluation recently (Taylor & Tyler, 2012), suggesting that the effects of efforts to identify and perhaps even improve effectiveness, varies for teachers and in systematic or patterned ways. Likewise, prior research on teacher effects (Hanushek & Rivkin, 2010) as well as effects of teacher evaluation (Taylor & Tyler, 2012) have long established differences in effects on reading and mathematics, with larger effects observed in mathematics in part because reading achievement may be more susceptible to out-of-school factors (e.g., at-home reading practices). This suggest that we should not expect similar effects of peers observers on teacher performance across subject areas. Thus, replication studies would be particularly useful in this latter effort and studies to determine how much collaboration explains the variance in outcomes from peer coaching as well as possible residual or secondary effects, to address the former. We commend those researchers who collected detailed field notes and even recordings of peer coaching sessions. Understanding the content and

---

\(^9\) We find it particularly interesting that peer coaching and feedback is utilized to a great extent in some countries, however, the literature included in our review did not align with these patterns, perhaps because this practice is widely used without adequate research or because our search criteria did not adequately illuminate the existing research.

\(^10\) One benefit to this is that studies have been conducted recently and thus, findings are more likely to generalize to today’s teachers, however, the frequency of research on this topic is limited.

\(^11\) In these two studies, peer coaching was not determined to have a negative impact, but positive outcomes were not explicitly noted.
quality of those peer coaching sessions would be valuable to illuminate why, perhaps, teachers report higher satisfaction when being evaluated by peers as opposed to their principals (Ford et al., 2018) as well as provide recommendations for future peer coaches. Furthermore, given the studies included in this review that indicated peer coaching is a positive form of professional development, it would be important to determine what characteristics of the peer coaching programs made them effective (see Garet, Porter, Desimone, Birman, & Yoon, 2001 for a review of the characteristics of effective professional development) and whether or not all peer coaching programs are equally effective. In a call for more quasi-experimental studies that measure change in instructional practices (among other important outcomes), it would be useful for future research to examine whether or not peer coaching models are more, less, or equally effective at improving instructional practice than other means (other professional development opportunities). Measuring whether or not changes in instructional practice translate to gains in student achievement (and other important student outcomes) would help extend this body of literature in important ways. Finally, studies with larger sample sizes as well as longitudinal studies need to be done in order to better represent the direction and size of peer coaching effects. Cost analysis would be particularly valuable for inclusion in such studies, because school leaders must make data-based decisions to maximize teaching and learning outcomes, but within a given budget (Hollands & Levin, 2017). Until further research is done, peer coaching cannot be applied in meaningful ways outside of the samples, conditions, and contexts in which they have already been studied.12

Practice. We are aware that the research included in the above review is not representative of peer feedback practices, globally (OECD, 2014). However, this discrepancy illuminates that peer coaching is already practiced widely in some settings. With that in mind, there are rich opportunities to study existing models of peer coaching, the implementation and sustainability process, and create powerful school-university partnerships to do so. For practitioners considering the implementation of a peer coaching program and limited evidence of the effect and cost of peer coaching programs, it would be useful to first conduct a small pilot. Building buy-in and trust are crucial, but districts and schools might want to test out various peer coaching models that vary based on whether or not peer coaches are chosen or matched (randomly or intentionally), length, and focus. However, we acknowledge this may be costly given all the other demands that principals and teachers face on a daily basis.

Policy. Given the potential of peer coaching, it would be useful for policies to help alleviate the barriers that practitioners and researchers face as it pertains to both the study and implementation of peer coaching. For example, policy might support the development and implementation of pilot studies, cost-benefit analyses, and offer resources, support, as well as funds to account for release time to engage in a deliberate study of peer coaching—both what it has been and what it could be. Given the overwhelming benefit of collaboration (which likely benefits teachers in a variety of ways), policies should encourage teacher engagement in collaborative opportunities that have the potential to improve their effectiveness and in teacher-as-researcher opportunities which allow teachers to play an active role in helping us better understand what works in improving their instructional practice and growth and development as a teacher and why.

Conclusion

12 In alignment with our earlier recommendations, we would strongly recommend replication studies even prior to the application of peer coaching in the same conditions, contexts, and with the same samples and peer coaching models as in the studies included in this review.
In conclusion, although the findings from our review primarily have implications for research (and much more of it), first, we see many opportunities to “close the research-practice-policy circle”. For example, in alignment with the call to scholars made by Holloway et al. (2017), we advocate for research that sits at the intersection of practice and policy and examines the function of peer coaching and feedback in different regulatory settings. Scholars might examine the perceptions and effectiveness of peer coaching and feedback in low-stakes as opposed to high-stakes teacher evaluation models (with the assumption that perhaps high-stakes teacher evaluation models reduce the effects of peer coaching on instructional improvement), as well as leveraging peers in both formative and summative ways (Ford et al., 2018). Furthermore, we agree with Dee and Wyckoff (2015) that any teacher evaluation system, procedure, and process inherently has error. Considering the practice of peer feedback, we might consider that peers could possibly provide misleading or inaccurate feedback, or that providing any feedback (regardless of its quality or quantity) diminishes teachers’ improvement efforts rather than enhances them. Furthermore, with any teacher evaluation system and its respective elements, districts must make choices in the context of various trade-offs (Dee & Wyckoff, 2015). Future research might explore what these trade-offs are when it comes to using peer observers (see Taylor & Tyler, 2012 for a notable example of such an analysis). For example, what is the cost of taking a teacher out of the classroom to serve as a peer observer? What is the cost of a school leader not doing the majority of classroom observations and feedback? It would also be important explore if and to what extent peer observers help address the limitations of using primarily school leaders as observers. For example, peer observers likely have a better knowledge of the day-to-day experiences of teachers, but do they use that knowledge in observing and providing feedback and if so, does it make a difference in how teachers perceive feedback and use it to change their practice? If leveraging peer observers frees up time for school leaders, how do they use this new found time and does it improve their effectiveness? Such research would help create a more synergetic approach to reducing the gap between research, practice, and policy and would promote a more deliberate and nuanced understanding of if and how peer coaching can be integrated into teacher evaluation in ways that help prioritize improving instructional practice and districts’ data-based decision-making.

Acknowledgements

We are grateful for anonymous reviewers who provided feedback and insights on earlier versions of this manuscript.

References *

* Denotes references that were included in the review of literature.


Improving Instructional Practice through Peer Observation and Feedback


Improving Instructional Practice through Peer Observation and Feedback


### About the Authors

**Brady L. Ridge**  
Utah State University  
brady.ridge@gmail.com

Brady Ridge is a doctoral student in instructional leadership at Utah State University. His research interests focus on principal feedback and teacher self-efficacy.

**Alyson L. Lavigne**  
Utah State University  
alyson.lavigne@usu.edu

Alyson Lavigne is an Assistant Professor of Instructional Leadership at Utah State University. Using her training as an educational psychologist and classroom researcher, her work explores teacher supervision, evaluation, and retention, and student, teacher, and leader dynamics and practices, particularly in schools that serve linguistically and culturally diverse learners.

### About the Guest Editor

**Audrey Amrein-Beardsley**  
Arizona State University  
audrey.beardsley@asu.edu
Audrey Amrein-Beardsley, PhD., is a Professor in the Mary Lou Fulton Teachers College at Arizona State University. Her research focuses on the use of value-added models (VAMs) in and across states before and since the passage of the Every Student Succeeds Act (ESSA). More specifically, she is conducting validation studies on multiple system components, as well as serving as an expert witness in many legal cases surrounding the (mis)use of VAM-based output.
archivo analíticos de políticas educativas
cONSEJO EDITORIAL

Editor Consultor: Gustavo E. Fischman (Arizona State University)
Editores Asociados: Felicitas Acosta (Universidad Nacional de General Sarmiento), Armando Alcántara Santuario (Universidad Nacional Autónoma de México), Ignacio Barrenechea, Jason Beech (Univerisidad de San Andrés), Angelica Buendia, (Metropolitan Autonomous University), Alejandra Falabella (Universidad Alberto Hurtado, Chile), Carmuxa Gómez-Bueno (Universidad de Granada), Veronica Gottau (Universidad Torcuato Di Tella), Carolina Guzmán-Valenzuela (Universidade de Chile), Antonio Lozano-Díaz (University of Almería), Antonio Luzon, (Universidad de Granada), Maria Teresa Martín Palomo (University of Almería), María Fernández Mellizo-Soto (University Complutense de Madrid), Tiburcio Moreno (Autonomous Metropolitan University-Cuajimalpa Unit), José Luis Ramírez, (Universidad de Sonora), Axel Rivas (Universidad de San Andrés), César Lorenzo Rodríguez Uribe (Universidad Marista de Guadalajara), Maria Veronica Santelices (Pontificia Universidad Católica de Chile)

Claudio Almonacid
Universidad Metropolitana de Ciencias de la Educación, Chile

Ana María García de Fanelli
Centro de Estudios de Estado y Sociedad (CEDES) CONICET, Argentina

Miriam Rodríguez Vargas
Universidad Autónoma de Tamaulipas, México

Miguel Ángel Arias Ortega
Universidad Autónoma de la Ciudad de México

Juan Carlos González Faraco
Universidad de Huelva, España

José Gregorio Rodríguez
Universidad Nacional de Colombia, Colombia

Xavier Besalú Costa
Universitat de Girona, España

María Clemente Linuesa
Universidad de Salamanca, España

Mario Rueda Beltrán Instituto de Investigaciones sobre la Universidad y la Educación, UNAM, México

Xavier Bonal Sarro Universidad Autónoma de Barcelona, España

Jaume Martínez Bonafé
Universitat de València, España

José Luis San Fabián Maroto
Universidad de Oviedo, España

Antonio Bolivar Boitia
Universidad de Granada, España

Alejandro Márquez Jiménez
Instituto de Investigaciones sobre la Universidad y la Educación, UNAM, México

Jurjo Torres Santomé, Universidad de la Coruña, España

José Joaquín Brunner Universidad Diego Portales, Chile

María Guadalupe Olivier Tellez, Universidad Pedagógica Nacional, México

Yengny Marisol Silva Laya
Universidad Iberoamericana, México

Damián Canales Sánchez Instituto Nacional para la Evaluación de la Educación, México

Miguel Pereyra Universidad de Granada, España

Ernesto Treviño Ronzón
Universidad Veracruzana, México

Gabriela de la Cruz Flores Universidad Nacional Autónoma de México

Mónica Pini Universidad Nacional de San Martín, Argentina

Ernesto Treviño Villarreal
Universidad Diego Portales Santiago, Chile

Marco Antonio Delgado Fuentes Universidad Iberoamericana, México

Omar Orlando Pulido Chaves Instituto para la Investigación Educativa y el Desarrollo Pedagógico (IDEP)

Antoni Verger Planells
Universidad Autónoma de Barcelona, España

Inés Dussel, DIE-CINVESTAV, México

José Ignacio Rivas Flores Universidad de Málaga, España

Catalina Wainerman
Universidad de San Andrés, Argentina

Pedro Flores Crespo Universidad Iberoamericana, México

Ernesto Treviño Villarreal
Universidad Diego Portales Santiago, Chile

Juan Carlos Yáñez Velazco
Universidad de Colima, México
arquivos analíticos de políticas educativas
conselho editorial

Editor Consultor: Gustavo E. Fischman (Arizona State University)
Editoras Associadas: Andréa Barbosa Gouveia (Universidade Federal do Paraná), Kaizo Iwakami Beltrao, (Brazilian School of Public and Private Management - EBAPE/FGVl), Sheizi Calheira de Freitas (Federal University of Bahia), Maria Margarida Machado, (Federal University of Goiás / Universidade Federal de Goiás), Gilberto José Miranda, (Universidade Federal de Uberlândia, Brazil), Marcia Pletsch, Sandra Regina Sales (Universidade Federal Rural do Rio de Janeiro)

Almerindo Afonso
Universidade do Minho
Portugal

Alexandre Fernandez Vaz
Universidade Federal de Santa Catarina, Brasil

José Augusto Pacheco
Universidade do Minho, Portugal

Rosanna Maria Barros Sá
Universidade do Algarve
Portugal

Regina Célia Linhares Hostins
Universidade do Vale do Itajaí, Brasil

Jane Paiva
Universidade do Estado do Rio de Janeiro, Brasil

Maria Helena Bonilla
Universidade Federal da Bahia
Brasil

Alfredo Macedo Gomes
Universidade Federal de Pernambuco
Brasil

Paulo Alberto Santos Vieira
Universidade do Estado de Mato Grosso, Brasil

Rosa Maria Bueno Fischer
Universidade Federal do Rio Grande do Sul, Brasil

Jefferson Mainardes
Universidade Estadual de Ponta Grossa, Brasil

Fabiany de Cássia Tavares Silva
Universidade Federal do Mato Grosso do Sul, Brasil

Alice Casimiro Lopes
Universidade do Estado do Rio de Janeiro, Brasil

Jader Janer Moreira Lopes
Universidade Federal Fluminense e Universidade Federal de Juiz de Fora, Brasil

António Teodoro
Universidade Lusófona
Portugal

Suzana Feldens Schwertner
Centro Universitário Univates
Brasil

Debora Nunes
Universidade Federal do Rio Grande do Norte, Brasil

Lillian do Valle
Universidade do Estado do Rio de Janeiro, Brasil

Geovana Mendonça Lunardi Mendes
Universidade do Estado de Santa Catarina

Alda Junqueira Marin
Pontifícia Universidade Católica de São Paulo, Brasil

Alfredo Veiga-Neto
Universidade Federal do Rio Grande do Sul, Brasil

Flávia Miller Naethe Motta
Universidade Federal Rural do Rio de Janeiro, Brasil

Dalila Andrade Oliveira
Universidade Federal de Minas Gerais, Brasil