

SPECIAL ISSUE
Policies and Practices of Promise
in Teacher Evaluation

education policy analysis
archives

A peer-reviewed, independent,
open access, multilingual journal



Arizona State University

Volume 28 Number 61

April 13, 2020

ISSN 1068-2341

**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

Journal website: <http://epaa.asu.edu/ojs/>
Facebook: /EPAAA
Twitter: @epaa_aape

Manuscript received: 9/3/2019
Revisions received: 2/1/2020
Accepted: 2/20/2020

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 observación 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 visitar 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 (Aronson, 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, 2020). 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)¹, as well as for those teachers who have remained (Dee & Wyckoff, 2015). Notably, these findings were

¹ 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)². 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 evaluations 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

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 (Horng, 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 feasible³.

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)⁴. 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 systematic⁵ 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”⁶).

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 1971⁷.

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 to 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, Fiegenger, & 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 1

Characteristics of Individual Literature

Author(s)	Document Type*	Theoretical Framework/Theorist(s)	Geographic Locale	Level	Context	Sample & Size	Definition of Peer Coaching	Measures	Study Design	Findings
Anderson, Barksdale, & Hite (2005)	J	None mentioned.	United States			34 elementary pre-service teachers	No clear definition; implicit is a peer-to-peer process	Weekly Journals, lesson plan, open-ended surveys, and observations	Qualitative	Students enjoyed their experience; observing and being observed.
Arnao, Kahrs & Kruskamp (2004)	J	None mentioned.	United States	Secondary		14 veteran teachers	"Peer coaching is a practice in which teachers work together through a process that includes preconferencing, observation, and postconferencing." (p. 26)	Interviews	Qualitative	Teachers reported the following reasons for participating in peer coaching: desire to learn, informal peer coaching experience, need for meaningful feedback, attraction to choices, dissatisfaction with traditional observation; meaningful feedback, self-directed learning, trust among peers, worth from extra work; increased morale in peer coaches.
Ben-Peretz, Gottlieb, & Gideon (2018)	J		Israel			2 in-service teachers	"Coaching is defined as an ongoing process in which experts engage with practice, with the purpose of continuous improvement." (p. 303)	Interviews	Qualitative	Teachers has positive perceptions of what collaboration can do to improve teaching.
Britton & Anderson (2010)	J	Development/R eflective Model of Teaching				4 Pre-service secondary teachers	No clear definition; implicit is a peer-to-peer	Open-ended interviews	Qualitative Program Evaluation & Action Research	Peer coaching was found to alter practices as reported by those who engaged in peer coaching. These findings suggest the importance of engaging pre-service teachers early in their program in peer coaching.
Bruce & Ross (2008)	J	Bandura	Canada	Primary		12 in-service teachers		Observations; Pre- and post-assessments	Mixed Method	First, teachers moved their practice toward standards-based methods; second, the professional development program had positive effects on teacher efficacy; and third, peer coaching caused participants to reflect more explicitly.
Castañeda-Londoño (2017)	J	None mentioned.	Colombia		Private Language School	3 in-service teachers	No clear definition; implicit is a peer-to-peer process	Interviews	Qualitative; Narrative	Before peer-coaching teachers were cautious concerning observation/feedback in the sense that although they can be used as a process to learn, the observer must be respectful of the expertise of the teacher being observed. Another perception was that of observation/feedback considered as a source of improvement, professionalism, and reflection about one's own actions through another's eyes, but could also be a threat to the teacher's professional identity. After using peer-coaching, teachers' perceptions included teacher identity tensions regarding self-image, intimidation, and self-improvement and the other that has to do with observation/feedback as means to understand one's own self as teacher.

Jao (2013)	J	Canada	Primary	4 in-service teachers	"...involves the pairing of two colleagues in a session with classroom observation, with feedback and discussion that allow both members to have a chance to learn from one another (Loucks-Horsley et al, 2003)."	Attitudes and beliefs surveys; interviews; observations	Qualitative	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.
Koch (2014)	D	United States: Pennsylvania	Secondary	17 in-service teachers	No explicit definition; implicit definition is reciprocal coaching	Quantitative scales	Quantitative: Regression	Teachers who participated in peer coaching experienced more collaboration, collegiality, and teaching self-efficacy.
Kohler, Crilley, Shearer, & Good (1997)	J	United States	Primary	4 elementary teachers	No clear definition; implicit is a peer-to-peer	Single case design	Quantitative	Their results indicated that areas that were not discussed during the seven coaching sessions were rarely implemented or refined in teachers' subsequent instructional practice illustrating the importance of intentional, systematic, frequent, and targeted coaching sessions in peer coaching models.
Kohler, Ezell, & Paluselli (1999)	J	United States	Primary	2 Kindergarten teachers	No explicit definition implicit definition is reciprocal coaching	Interviews	Mixed Method	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 concluded (as measured in the maintenance phase).
Lee & Choi (2013)	J	Korea		10 pre-service secondary physical education teachers	No clear definition; implicit is a peer-to-peer process	Journals, evaluation forms, interviews	Qualitative	Findings indicated that peer coaching facilitated reflection by helping teachers face barriers and connecting planning and instruction with reflection. Through reflections peer coaches were able to reiterate and accept their practices, confront and extend, and adjust and re-aim their instructional goals.
Licklider (1995)	J	United States	Secondary	11 volunteer secondary teachers	No explicit definition implicit definition is reciprocal coaching	Audiotapes for coaching sessions	Qualitative	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

Table 1, cont.

Ma, Xin, & Du (2018)	J	China	None mentioned.	20 in-service teachers (Treatment = 10; Control = 10)	"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).	Surveys; videos	Quantitative: Quasi-experimental	Learning participation and pedagogy was higher in treatment group, but design skills were equal.
Meng, Tajaroensak, & Scepho (2013)	J	China	None mentioned.	12 efl teachers and 105 second year english major students	No explicit definition; implicit definition is reciprocal coaching		Mixed Method	Significant gains in knowledge from pre to post, teachers found peer coaching to be helpful.
Murray, Ma, & Mazur (2009)	J	United States	Vygotsky; specifically zone of proximal development	Experimental group had 6 teachers and 202 students. Control group had 5 teachers and 105 students.		Quasi Experimental. Used surveys and post observation conference	Mixed Method	Peer coaching was not associated with any improvement in mathematics achievement of students.
Neubert & McAllister (1993)	J	United States	None mentioned.	144 pre-service teachers	No clear definition; implicit is a peer-to-peer	Questionnaires	Qualitative	93% loved or liked the peer coaching. There were a few problems: students were concerned about offending partners with polish suggestions; difficulty thinking about polish skills; difficulty identifying what to focus on for lesson; help students apply strategies. Reflections indicated ability to apply webbing, vocab placement, cooperative learning
Ovens (2004)	J	New Zealand	Not explicit	12 pre-service physical education teachers	"a process where student teachers work collaboratively with a partner whil on practicum to plan lessons, observe each other's teaching and provide feedback in post lesson discussions." (p. 47)	Surveys; observations; interviews	Action Research	1. Peer coaching practicum was perceived to be positive - increase theorizing, student thinking about teaching, provided students with more autonomy over decision making 2. Peer coaching practicum was problematic in that there was not enough time in school, high workloads, lack of skills to analyze lessons, disrupted or poor quality practicum settings, pressure to be a relief teacher.

Table 1, cont.

Pearce, De la Fuente, & Weinburg (2019)	J	Trust Theory	United States	Secondary	Biology	29 in-service teachers	"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 focuses on performance." (p. 117)	Focus groups and notes (pre and post); observations of peer coaching	Qualitative: Case Study	Attitudes about coaching affect outcomes of coaching. Many teachers attributed the coaching as the reason for increased teaching competence. Key difference is this was a two year coaching process and not just a one time thing.
Phillips & Glickman (1991)	J	Conceptual Systems Theory; Developmental Matching Model	United States; Georgia	Primary		22 in-service teachers	"a process in which classroom teachers observe one another teach, give feedback concerning the observation, and together develop an instructional improvement plan." (Phillips, 1989 as cited in Phillips and Glickman, p. 21)	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	Quantitative: Quasi-experimental Program Evaluation	1. Peer coaching resulted in a slight increase in number of interactions with other teachers (collaboration, reduced isolation) 2. 18 of 22 teachers planned to change their teaching. 3. Teachers found the following benefits of peer coaching: focus on specifics of teaching, gain new ideas, and develop new insight. 4. Peer coaching raised teachers' conceptual levels
Pollara (2012)	D		United States	Primary		15 in-service teachers		Surveys (pre and post); focus groups; interviews	Action Research	1. Peer coaching was related to a reported increase in repertoire of instructional skills, growth in classroom management, knowledge of the curriculum in grades other than the one currently assigned. 2. Frequency of the use of common planning time increased. 3. Peer coaching increased collaboration and reduced isolation.
Porras, Díaz, & Nieves (2018)	J	Professional Development, Reverse Mentoring, & Peer Coaching	Colombia	Primary		10 in-service elementary teachers	"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." (p. 175)	Survey; English proficiency test; field notes; interviews	Mixed Method	1. Reverse mentoring allowed inservice teachers to observe pre-service teachers and reflect on their own teaching and successful practices for teaching English to children. 2. Peer coaching provided support and collaborative around a shared activity where participants enhanced their knowledge and were able to test new teaching tools, techniques, and strategies
Prince, Snowden, & Matthews (2010)	J	None mentioned.	United Kingdom	Secondary	Science	38 pre-service teachers 3 dyads (6 teachers total with 3 special ed and 3 general ed)	No clear definition; implicit is a peer-to-peer process	Questionnaires to provide quant and qual data	Mixed Method	Improved their ability to respond to others in a professional capacity; developed listening skills, joint planning skills; developed confidence.
Scheeler, Congdon & Stansbery (2010)	J	None mentioned.	United States				No clear definition; implicit is a peer-to-peer	Bug-in-car (BIE) corrective feedback	Quantitative	Maintained behavior at 90% or higher after just three BIE sessions; maintained it post-intervention, and generalized the behavior to a different setting without the peer coach present

Table 1, cont.

Showers (1984)	R	None mentioned.	United States	21 teachers and 6 peer coaching in 2 districts	No clear definition; implicit is a peer-to-peer	Student data, lesson plans, observations, and interviews.	Qualitative	Peer coaches can be trained relatively quickly. Transfer of training was better for coached vs. uncoached teachers. Students of coached teachers scored higher on concept attainment.
Syh-Jong & Hsu-Chuan (2009)	J	Pedagogical Content Knowledge	Taiwan	4 in-service teachers and 123 secondary students	No clear definition; implicit is a peer-to-peer	Student Surveys, teacher surveys, reflective journals, and interviews	Mixed Method	Teachers gained greater confidence and student and teacher reflections were important for increasing pedagogical content knowledge through peer coaching and feedback.
Shui-Fong & Wing-Shuen (2008)	J	None mentioned.	Hong Kong	355 in-service teachers across 2 studies	"Peer coaching is the confidential process through which two or more professional colleagues work together to reflect on current practices; expand, refine, and build new skills, share ideas; teach one another; conduct classroom research; or solve problems in the workplace." (p. 68)	Surveys measuring teachers' acceptance of peer coaching, perception of collegial school culture, and goal orientation.	Quantitative: Correlational	1. Teachers were more likely to participate in and accept peer coaching if they reported higher collegiality in their schools and preferred learning goals.
Slater & Simmons (2001)	J	None mentioned.	United States: Texas	80 in-service teachers		Evaluative: Teacher Interview Questions, Participation Surveys, Peer Coaching Evaluation	Qualitative	Helped teachers overcome isolation and helped enhance self-reported teaching skills.
Sparks (1986)	J	None mentioned.	United States	19 junior high teachers	No clear definition; implicit is a peer-to-peer	Peer observation and feedback and training/coaching sessions from coach	Qualitative	Peer observation and feedback appeared to result in greatest improvements, however, teachers were not randomly assigned to condition.
Thijs, & Van den Berg (2002)	J	Vygotskian, specifically teachers' zone of proximal development as a guide for enacting peer coaching	Botswana	120 in-service teachers	"a confidential relationship between professional colleagues working together to reflect on their teaching and share ideas in order to improve their professional skills." (p. 55)	Observations of in-service course to assess implementation fidelity; survey of teachers' opinions about the usefulness of course; survey on attitudes towards peer coaching (pre and post); interviews, recorded coaching sessions, teacher logs	Mixed method	1. In-service course positively affected participants' attitudes towards peer coaching. 2. Curriculum materials as part of the PD program was effective for changing practices for mathematics teachers, but not physics teachers. 3. Teachers considered peer coaching to be beneficial, including the following benefits: new insights and suggestions for improvement

Table 1, cont.

Vacilotto & Cummings (2007)	J	None mentioned.	Brazil	16 graduate student pre-service teachers	No explicit definition implicit definition is reciprocal coaching	Peer lesson plans, peer observation, reflective journals, pre and post surveys, and post interview	Findings indicate that peer coaching facilitated exchange of teaching methods and materials, fostered development of teaching skills, and made participants rethink their own teaching methods and styles. The study also revealed which behaviours participants thought were most effective for supporting a successful relationship among peers in a peer coaching programme.	Qualitative
Wynn & Kromrey (1999)	J	None mentioned.	United States	81 pre-service teachers 78 continued in second year follow-up	"a training method in which pairs of practicum students, student teachers, or classroom teachers observe each other and provide consultative assistance in correctly applying teaching skills and proposing alternative solutions to recognized instructional needs" (p. 22)	Peer coaching forms; surveys on advantages and disadvantages of peer coaching; weekly dialogue journals; audiotapes and videotapes; informal interviews	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 practicum student and peer coach.	Mixed Method
Wynn & Kromrey (2000)	J	None mentioned.	United States	26 pre-service teachers	"a training method in which pairs of practicum students or interns observe each other and provide consultative assistance in correctly implementing teaching strategies and proposing alternative solutions to recognized instructional concerns." (p. 74)	Students' observations; Peer coach and practicum student feedback; students' reflections on their teaching experiences	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 towards 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.	Qualitative

Table 1, cont.

Zwart, Wubbels, Bergen, & Bolhuis (2007)	J	None mentioned.	Netherlands	28 in-service teachers which (14 dyads)	No explicit definition implicit definition is reciprocal coaching diaries	Surveys (pre and post); digital diaries	Mixed Method	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.
Zwart, Wubbels, Bergen, & Bolhuis (2009)	J	None mentioned.	Netherlands	28 in-service teachers (14 dyads)	No explicit definition implicit definition is reciprocal coaching diaries	Surveys (pre and post); digital diaries	Mixed Method	Peer coaching was related to moderate growth with teacher learning.

* 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, Tajarosuk, & 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 review⁸, it is possible that we failed to include studies that would

⁸ 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 research⁹, 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 young¹⁰, 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, 2009¹¹) 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 studied¹².

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 to 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.

- Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics*, 25(1), 95–135.
<https://doi.org/10.1086/508733>
- Ackland, R. (1991) A review of the peer coaching literature. *Journal of Staff Development*, 12(1), 22–26.
- Adnot, M., Dee, T., Katz, V., & Wyckoff, J. (2017). Teacher turnover, teacher quality, and student achievement. *Educational Evaluation and Policy Analysis*, 39, 54–76.
<https://doi.org/10.3102/0162373716663646>

- Allen, D.W., & LeBlanc, A.C. (2005). *Collaborative peer coaching that improves instruction: The 2 + 2 performance appraisal model*. Corwin Press.
- American Educational Research Association. (2009). Standards for reporting on humanities-oriented research in AERA publications. *Educational Researcher*, 36(6), 481–486.
<https://doi.org/10.3102/0013189X09341833>
- American Psychological Association. (2020). *Publication manual of the American Psychological Association* (7th ed.). Author.
- *Anderson, N. A., Barksdale, M. A., & Hite, C. E. (2005). Preservice teachers' observations of cooperating teachers and peers while participating in an early field experience. *Teacher Education Quarterly*, 32(4), 97–117.
- *Arnau, L., Kahrs, J., & Kruskamp, B. (2004). Peer coaching: Veteran high school teachers take the lead on learning. *NASSP Bulletin*, 88(639), 26–41.
<https://doi.org/10.1177/019263650408863904>
- Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H.F....Shepard, L. A. (2010). *Problems with the use of student test scores to evaluate teachers*. [Briefing Paper #278]. Economic Policy Institute. Retrieved from <https://www.epi.org/publication/bp278/>
- Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology*, 52, 1–26. <https://doi.org/10.1146/annurev.psych.52.1.1>
- *Ben-Peretz, M., Gottlieb, E., & Gideon, I. (2018). Coaching between experts – opportunities for teachers' professional development. *Teacher Development*, 22(3), 303–313.
<https://doi.org/10.1080/13664530.2018.1438310>
- *Britton, L. R., & Anderson, K. A. (2010). Peer coaching and pre-service teachers: Examining an underutilized concept. *Teaching and Teacher Education: An International Journal of Research and Studies*, 26(2), 306–314. <https://doi.org/10.1016/j.tate.2009.03.008>
- Brophy, J., & Good, T. (1986). Teacher behavior and student achievement. In M. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 328–375). Macmillan.
- *Bruce, C. D., & Ross, J. A. (2008). A model for increasing reform implementation and teacher efficacy: Teacher peer coaching in grades 3 and 6 mathematics. *Canadian Journal of Education*, 31(2), 346–370.
- Burgess, S., Shenila, R., Taylor, E. S. (2019). *Teacher peer observation and student test scores: Evidence from a field experiment in English secondary schools*. Working paper. Retrieved from <https://scholar.harvard.edu/files/erictaylor/files/teacher-peer-obsv-brt-jan-19.pdf>
- *Castañeda-Londoño, A. (2017). Exploring 23 research teachers' perceptions about peer-coaching as a professional development activity of knowledge construction. *HOW*, 24(2), 80–101.
<https://doi.org/10.19183/how.24.2.345>
- Cherasaro, T. L., Brodersen, R. M., Reale, M. L., & Yanoski, D. C. (2016). *Teachers' responses to feedback from evaluators: What feedback characteristics matter?* (REL 2017-190). Regional Educational Laboratory Central.
- Croft, M., Guffy, G., & Vitale, D. (2018, July). *The shrinking use of growth: Teacher evaluation legislation since ESSA*. Retrieved from <https://www.act.org/content/dam/act/unsecured/documents/teacher-evaluation-legislation-since-essa.pdf>
- Dee, T. S., & Wyckoff, J. (2015). Incentives, selection, and teacher performance: Evidence from IMPACT. *Journal of Policy Analysis and Management*, 34, 267–297.
Doi:10.1002/pam.21818
- Dimitriadis, G. (2009). Series editors introduction. In J. Anyon (Ed.), *Theory and educational research: Toward critical social explanation* (pp. vii–ix). Routledge.

- Donaldson, M. L., & Woulfin, S. (2018). From tinkering to going “rouge”: How principals use agency when enacting new teacher evaluation systems. *Educational Evaluation and Policy Analysis*, 40(4), 531–556. <https://doi.org/10.3102/0162373718784205>
- Flores, M. A., & Derrington, M. L. (2017). School principals’ views of teacher evaluation policy: Lessons learned from two empirical studies. *International Journal of Leadership in Education*, 20(4), 416–431. <https://doi.org/10.1080/13603124.2015.1094144>
- Ford, T., Lavigne, A. L., Fiegner, A., & Si, S. (2020). Understanding district support for leader development and success in the accountability era: A review of literature using social-cognitive theories of motivation. *Review of Educational Research*. Advance online publication. <https://doi.org/10.3102/0034654319899723>
- Ford, T. G., Urick, A., & Wilson, A. (2018). Exploring the effect of supportive teacher evaluation experiences on U.S. teachers’ job satisfaction. *Educational Policy Analysis Archives*, 26(59). <http://dx.doi.org/10.14507/epaa.26.3559>
- Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. W. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, 38(4), 915–945. <https://doi.org/10.3102/00028312038004915>
- Glanz, J., & Hazi, H. M. (2019). Shedding light on the phenomenon of supervision traveling incognito: A field’s struggles for visibility. *Journal of Educational Supervision*, 2(1), 1–21. <https://doi.org/10.31045/jes.2.1.1>
- Glickman, C. D., Gordon, S. P., & Ross-Gordon, J. M. (2017). *SuperVision and instructional leadership* (10th ed.). Pearson.
- Goldhaber, D., & Brewer, D. (1999). Teacher licensing and student achievement. In C. Finn & M. Kanstoroom (Eds.), *Better teachers, better schools* (pp. 83–102). Thomas B. Fordham Institute.
- Goldring, E. B., Grissom, J. A., Rubin, M., Neumerski, C. M., Cannata, M., Drake, T., & Schuermann, P. (2015). Make room value added: Principals’ human capital decisions and the emergence of teacher observation data. *Educational Researcher*, 44(2), 96–104. <https://doi.org/10.3102/0013189X15575031>
- *Gonen, S. I. K. (2016). A study on reflective reciprocal peer coaching for pre-service teachers: Change in reflectivity. *Journal of Education and Training Studies*, 4(7), 211–225. <https://doi.org/10.11114/jets.v4i7.1452>
- Good, T., Biddle, B., & Brophy, J. (1975). *Teachers make a difference*. Holt, Rinehart, and Winston.
- Greenhalgh, T., & Peacock, R. (2005). Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources. *BMJ*, 331(7524), 1064–1065. <https://doi.org/10.1136/bmj.38636.593461.68>
- *Hall, L. & McKeen, R. (1989). Increased professionalism in the work environment of teachers through peer coaching. *Education*, 109(3), 310–317.
- *Hall, L., & McKeen, R. L. (1991). Peer coaching as an organization development intervention in the public schools. *Education*, 111(4), 553–558.
- Hanushek, E. A., & Rivkin, S. G. (2010). Generalizations about using value-added measures of teacher quality. *American Education Review*, 100(2), 267–271. <https://doi.org/10.1257/aer.100.2.267>
- *Hasbrouck, J. E., & Christen, M. H. (1997). Providing peer coaching in inclusive classrooms: A tool for consulting teachers. *Intervention in School and Clinic*, 32, 172–177. <https://doi.org/10.1177/105345129703200308>
- Hazi, H. M. (1994). Teacher evaluation-supervision dilemma: A case of entanglements and irreconcilable differences. *Journal of Curriculum and Supervision*, 9(2), 195–216.

- Hazi, H. M., & Rucinski, D. A. (2019). Teacher evaluation as a policy target for improved student learning: A fifty-state review of statute and regulatory action since NCLB. *Education Policy Analysis Archives*, 17(5), 1–21. <https://doi.org/10.14507/epaa.v17n5.2009>
- Hill, H., Charalambous, C. Y., & Kraft, M. A. (2012). When interrater-reliability is not enough: Teacher observation systems and a case for the generalizability theory. *Educational Researcher*, 41, 56–64.
- Ho, A. D., & Kane, T. J. (2013). *The reliability of classroom observations by school personnel*. Bill & Melinda Gates Foundation.
- Hollands, F. M., & Levin, H. M. (2017). *The critical importance of costs for education decisions* (REL 2017–274). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Analytic Technical Assistance and Development. Retrieved from <http://ies.ed.gov/ncee/edlabs>.
- Holloway, J., Sørensen, T. B., & Verger, A. (Eds.)(2017). Global perspectives on high-stakes accountability [special issue]. *Education Policy Analysis Archives*, 25(85).
- Holloway, J., Sørensen, T. B., & Verger, A. (2017). Global perspectives on high-stakes accountability [special issue]: An introduction *Education Policy Analysis Archives*, 25(85), 1–18. <http://dx.doi.org/10.14507/epaa.25.3325>
- Howell, W. G. (2015). Results of President Obama’s Race to the Top. *Education Next*, 15(4), 58–66. Retrieved from <http://educationnext.org/results-president-obama-race-to-the-top-reform/>
- *Jao, L. (2013). Peer coaching as a model for professional development in the elementary mathematics context: Challenges, needs and rewards. *Policy Futures in Education*, 11(3), 290–297. <https://doi.org/10.2304/pfie.2013.11.3.290>
- Joyce, B., & Showers, B. (1980). Improving inservice training: The messages of research. *Educational Leadership*, 37, 379–385.
- Joyce, B., & Showers, B. (1981, April). *Teacher training research: Working hypotheses for program design and directions for further study*. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles.
- Joyce, B., & Showers, B. (1982). The coaching of teaching. *Educational Leadership*, 40(1), 4–8.
- Joyce, B., & Showers, B. (1985). *Student achievement through staff development: Fundamentals of school renewal* (2nd ed.). Longman.
- Joyce, B., & Showers, B. (1987). The low cost arrangements for peer coaching. *Journal of Staff Development*, 8(1), 22–28.
- Joyce, B., & Showers, B. (1988). *Student achievement through self development* (1st ed.). Longman.
- Joyce, B., & Showers, B. (2002). *Student achievement through self development* (3rd ed.). Association for Supervision and Curriculum Development.
- Kersten, T. A., & Israel, M. S. (2005). Teacher evaluation: Principals’ insights and suggestions for improvement. *Planning and Changing*, 36(1–2), 47–67.
- *Koch, M.-A. (2014, January 1). The relationship between peer coaching, collaboration and collegiality, teacher effectiveness and leadership (doctoral dissertation). Walden University.
- *Kohler, F. W., Crilley, K. M., Shearer, D. D., & Good, G. (1997). Effects of peer coaching on teacher and student outcomes. *Journal of Educational Research*, 90(4), 240–250. <https://doi.org/10.1080/00220671.1997.10544578>
- *Kohler, F. W., Ezell, H. K., & Paluselli, M. (1999). Promoting changes in teachers’ conduct of student pair activities: An examination of peer coaching. *The Journal of Special Education*, 33(3), 154–165.

- Kraft, M. A., & Gilmour, A. F. (2016). Can principals promote teacher development as evaluators? A case study of principals' views and experiences. *Educational Administration Quarterly*, 52(5), 711–753. <https://doi.org/10.1177/0013161X16653445>
- Lavigne, A. L., & Chamberlain, R. (2017). Teacher evaluation in Illinois: School leaders' perceptions and practices. *Educational Assessment, Evaluation and Accountability*, 29(2), 179–209. <https://doi.org/10.1007/s11092-016-9250-0>
- Lavigne, A. L., & Good, T. (2019). *Enhancing teacher education, development, and evaluation: Lessons learned from educational reform*. Routledge.
- Lavigne, A. L., & Good, T. (2020, January 13). *Addressing teacher evaluation appropriately: A research brief for policymakers*. Division 15: American Psychological Association. [EdPsych.us/AddressingTeacherEvaluation](https://www.edpsych.org/AddressingTeacherEvaluation)
- Lavigne, A. L., Good, T. L., & Marx, R. W. (Eds). (2014). High-stakes teacher evaluation: High cost—big losses [Special issue]. *Teachers College Record*, 116(1).
- *Lee, O., & Choi, E. (2013). Utilizing peer coaching to facilitate pre-service physical education teachers' reflection. *The Asia-Pacific Education Reseracher*, 22(2), 147–154. <https://doi.org/10.1007/s40299-012-0007-3>
- *Licklider, B. L. (1995) The effects of peer coaching cycles on teacher use of complex teaching skill and teacher's sense of efficacy. *Journal of Personnel Evaluation in Education*, 9, 55–68. <https://doi.org/10.1007/BF00975249>
- Lu, H. L. (2010). Research on peer coaching in preservice teacher education – A review of literature. *Teaching and Teacher Education*, 26(4), 748–753. <https://doi.org/10.1016/j.tate.2009.10.015>
- *Ma, N., Xin, S., & Du, J-Y. (2018). A peer coaching-based professional development approach to improving the learning participation and learning design skills of in-service teachers. *Journal of Educational Technology & Society*, 21(2), 291–304.
- *Meng, J., Tajaroensuk, S., & Seepho, S. (2013). The multilayered peer coaching model and the in-service professional development of tertiary EFL teachers. *International Education Studies*, 6(7), 18–31. <https://doi.org/10.5539/ies.v6n7p18>
- *Murray, S., Ma, X., & Mazur, J. (2009). Effects of peer coaching on teachers' collaborative interactions and students' mathematics achievement. *Journal of Educational Research*, 102(3), 203–212. <https://doi.org/10.3200/JOER.102.3.203-212>
- National Council on Teacher Quality. (2017). *Running in place: How new teacher evaluations fail to live up to the promises*. Retrieved from http://www.nctq.org/dmsView/Final_Evaluation_Paper
- *Neubert, G. A., & McAllister, E. A. (1993). Peer coaching in preservice education. *Teacher Education Quarterly*, 20(4), 77–84.
- Organisation for Economic Co-operation and Development [OECD]. (2014a). *TALIS 2013 results: An international perspective on teaching and learning*. Retrieved from http://www.keepeek.com/Digital-Asset-Management/oecd/education/talis-2013-results_9789264196261-en
- Organisation for Economic Co-operation and Development [OECD]. (2014b). Results from TALIS 2013: United States of America <http://www.oecd.org/unitedstates/TALIS2013-country-note-US.pdf>.
- *Ovens, A. (2004). Using peer coaching and action research to structure the practicum: An analysis of student teacher perceptions. *Journal of Physical Education New Zealand*, 37, 45–60.
- Papay, J. P., Taylor, E. S., Tyler, J. H., & Laski, M. (2016). *Learning job skills from colleagues at work: Evidence from a field experiment using teacher performance data* (NBER Paper No. 21986). Cambridge, MA: National Bureau of Economic Research. <https://doi.org/10.3386/w21986>

- *Pearce, E., de la Fuente, Y., Hartweg, B., & Weinburgh, M. (2019). Peer-Coaching as a component of a professional development model for biology teachers. *School Science and Mathematics, 119*(3), 117–126. <https://doi.org/10.1111/ssm.12326>
- *Phillips, M. D., & Glickman, C. D. (1991). Peer coaching: Developmental approach to enhancing teacher thinking. *Journal of Staff Development, 12*(2), 20–25.
- *Pollara, J. (2012, January 1). *Peer coaching: Teachers as leaders, teachers as learners*.
- *Porrás, N. I., Díaz, L. S., & Nieves, M. M. (2018). Reverse mentoring and peer coaching as professional development strategies. *Colombian Applied Linguistics Journal, 20*(2), 165–179. <https://doi.org/10.14483/22487085.12422>
- *Prince, T., Snowden, E., & Matthews, B. (2010). Utilising peer coaching as a tool to improve student-teacher confidence and support the development of classroom practice. *Literacy Information and Computer Education Journal, 1*(1), 49–51. <https://doi.org/10.20533/licej.2040.2589.2010.0007>
- Robbins, P. (2015). *Peer coaching to enrich professional practice, school culture, and student learning*.
- Rubie-Davies, C. (2014). *Becoming a high expectation teacher: Raising the bar*. Routledge.
- *Scheeler, M. C., Congdon, M., & Stansbery, S. (2010). Providing immediate feedback to coteachers through bug-in-ear technology: An effective method of peer coaching in inclusion classrooms. *Teacher Education and Special Education, 33*(1), 83–96. <https://doi.org/10.1177/0888406409357013>
- *Showers, B. (1984). *Peer coaching: A strategy for facilitating transfer of training*. Report to the U.S. Department of Education. Center for Educational Policy and Management, University of Oregon. ERIC EC 271 849.
- Showers, B. (1985). Teachers coaching teachers. *Educational Leadership, 42*, 43–48.
- Showers, B., & Joyce, B. (1996). The evolution of peer coaching. *Educational Leadership, 53*, 12–16.
- *Shui-Fong, L., & Wing-Shuen, L. (2008). Teachers' acceptance of peer coaching: Impact of collegiality and goal orientation. *Journal of School Connections, 1*(1), 3–24.
- *Slater, C. L., & Simmons, D. L. (2001). The design and implementation of a peer coaching program. *American Secondary Education, 29*(3), 67–76.
- Snodgrass-Rangel, V. (2018). A review of the literature on principal turnover. *Review of Educational Research, 88*(1), 87–124. <https://doi.org/10.3102/0034654317743197>
- *Sparks, G. M. (1986). The effectiveness of alternative training activities in changing teaching practices. *American Educational Research Journal, 23*, 217–225. <https://doi.org/10.3102/00028312023002217>
- Stecher B. M., Holtzman, D. J., Garet, M. S., Hamilton, L. S., Engberg, J...Chambers, J. (2018). *Improving teaching effectiveness. Final report. The intensive partnerships for effective teaching through 2015–2016*. <https://doi.org/10.7249/RR2242>
- *Syh-Jong, J., & Hsiu-Chuan, S. (2009). Developing in-service science teachers' PCK through a peer coaching-based model. *Journal of Education Research, 3*(1, 2), 87–108.
- Taylor, E. S., & Tyler, J. H. (2012). The effect of evaluation on teacher performance. *American Economic Review, 102*(7), 3628–3651. doi:10.1257/aer.102.7.3628
- Teddlie, C., & Reynolds, D. (2000). *The international handbook of school effectiveness research*. Falmer Press.
- *Thijs, A., & van den Berg, E. (2002). Peer coaching as part of a professional development program for science teachers in Botswana. *International Journal of Educational Development, 22*(1), 55–68. [https://doi.org/10.1016/S0738-0593\(00\)00078-X](https://doi.org/10.1016/S0738-0593(00)00078-X)
- Trujillo, T. (2013). The reincarnation of the effective schools research: Rethinking the literature on district effectiveness. *Journal of Educational Administration, 51*, 426–452.

- Trujillo, T. (2016). Learning from the past to chart new directions in the study of school district effectiveness. In A. J. Daly & K. S. Finnegan (Eds.), *Thinking and acting systematically: Improving school districts under pressure* (pp. 11-47). American Educational Research Association. doi:10.3102/978-0-935302-46-2_1
- *Vacilotto, S., & Cummings, R. (2007). Peer coaching in TEFL/TESL programmes. *ELT Journal*, 61(2), 153–160. <https://doi.org/10.1093/elt/ccm008>
- Van der Lans, R. M., van de Grift, W. J. C. M., van Veen, K., & Fokkens-Bruinsma, M. (2016). Once is not enough: Establishing reliability criteria for feedback and evaluation decisions based on classroom observations. *Studies in Educational Evaluation*, 50, 88–95.
- Wohlin, C. (2014). Guidelines for snowballing and systematic literature studies and a replication in software engineering. *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering, Article 18*, 1–10. <https://doi.org/10.1016/j.stueduc.2016.08.001>
- *Wynn, M. J., & Kromrey, J. D. (1999). Paired peer placement with peer coaching in early field experiences: results of a four-year study. *Teacher Education Quarterly*, 26(1), 21–38.
- *Wynn, M., & Kromrey, J. (2000). Paired peer placement with peer coaching to enhance prospective teachers' professional growth in early field experience. *Action in Teacher Education*, 22(2), 73–83.
- *Zwart, R. C., Wubbels, T., Bergen, T. C. M., & Bolhuis, S. (2007). Experienced teacher learning within the context of reciprocal peer coaching. *Teachers and Teaching: Theory and Practice*, 13(2), 165–187. <https://doi.org/10.1080/13540600601152520>
- *Zwart, R. C., Wubbels, T., Bergen, T., & Bolhuis, S. (2009). Which characteristics of a reciprocal peer coaching context affect teacher learning as perceived by teachers and their students? *Journal of Teacher Education*, 60(3), 243–257. <https://doi.org/10.1177/0022487109336968>

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.

Special Issue Policies and Practices of Promise in Teacher Evaluation

education policy analysis archives

Volume 28 Number 61

April 13, 2020

ISSN 1068-2341



Readers are free to copy, display, distribute, and adapt this article, as long as the work is attributed to the author(s) and **Education Policy Analysis Archives**, the changes are identified, and the same license applies to the derivative work. More details of this Creative Commons license are available at <https://creativecommons.org/licenses/by-sa/2.0/>. **EPAA** is published by the Mary Lou Fulton Institute and Graduate School of Education at Arizona State University. Articles are indexed in CIRC (Clasificación Integrada de Revistas Científicas, Spain), DIALNET (Spain), [Directory of Open Access Journals](#), EBSCO Education Research Complete, ERIC, Education Full Text (H.W. Wilson), QUALIS A1 (Brazil), SCImago Journal Rank, SCOPUS, SOCOLAR (China).

Please send errata notes to Audrey Amrein-Beardsley at audrey.beardsley@asu.edu

Join **EPAA's Facebook community** at <https://www.facebook.com/EPAAAPE> and **Twitter feed** @epaa_aape.

education policy analysis archives
editorial board

Lead Editor: **Audrey Amrein-Beardsley** (Arizona State University)

Editor Consultor: **Gustavo E. Fischman** (Arizona State University)

Associate Editors: **Melanie Bertrand, David Carlson, Lauren Harris, Eugene Judson, Mirka Koro-Ljungberg, Daniel Liou, Scott Marley, Molly Ott, Iveta Silova** (Arizona State University)

Cristina Alfaro
San Diego State University

Gary Anderson
New York University

Michael W. Apple
University of Wisconsin, Madison

Jeff Bale
University of Toronto, Canada
Aaron Bevenot SUNY Albany

David C. Berliner
Arizona State University
Henry Braun Boston College

Casey Cobb
University of Connecticut

Arnold Danzig
San Jose State University
Linda Darling-Hammond
Stanford University

Elizabeth H. DeBray
University of Georgia

David E. DeMatthews
University of Texas at Austin

Chad d'Entremont Rennie Center
for Education Research & Policy

John Diamond
University of Wisconsin, Madison

Matthew Di Carlo
Albert Shanker Institute

Sherman Dorn
Arizona State University

Michael J. Dumas
University of California, Berkeley

Kathy Escamilla
University of Colorado, Boulder

Yariv Feniger Ben-Gurion
University of the Negev

Melissa Lynn Freeman
Adams State College

Rachael Gabriel
University of Connecticut

Amy Garrett Dikkers University
of North Carolina, Wilmington

Gene V Glass
Arizona State University

Ronald Glass University of
California, Santa Cruz

Jacob P. K. Gross
University of Louisville
Eric M. Haas WestEd

Julian Vasquez Heilig California
State University, Sacramento
Kimberly Kappler Hewitt
University of North Carolina
Greensboro

Aimee Howley Ohio University

Steve Klees University of Maryland

Jaekyung Lee SUNY Buffalo

Jessica Nina Lester

Indiana University
Amanda E. Lewis University of
Illinois, Chicago

Chad R. Lochmiller Indiana
University

Christopher Lubienski Indiana
University

Sarah Lubienski Indiana University

William J. Mathis
University of Colorado, Boulder

Michele S. Moses
University of Colorado, Boulder

Julianne Moss
Deakin University, Australia

Sharon Nichols
University of Texas, San Antonio

Eric Parsons
University of Missouri-Columbia

Amanda U. Potterton
University of Kentucky

Susan L. Robertson
Bristol University

Gloria M. Rodriguez
University of California, Davis

R. Anthony Rolle
University of Houston

A. G. Rud
Washington State University

Patricia Sánchez University of
University of Texas, San Antonio

Janelle Scott University of
California, Berkeley

Jack Schneider University of
Massachusetts Lowell

Noah Sobe Loyola University

Nelly P. Stromquist
University of Maryland

Benjamin Superfine
University of Illinois, Chicago

Adai Tefera
Virginia Commonwealth University

A. Chris Torres
Michigan State University

Tina Trujillo
University of California, Berkeley

Federico R. Waitoller
University of Illinois, Chicago

Larisa Warhol
University of Connecticut

John Weathers University of
Colorado, Colorado Springs

Kevin Welner
University of Colorado, Boulder

Terrence G. Wiley
Center for Applied Linguistics

John Willinsky
Stanford University

Jennifer R. Wolgemuth
University of South Florida

Kyo Yamashiro
Claremont Graduate University

Miri Yemini
Tel Aviv University, Israel

archivos 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** (Universidad de San Andrés), **Angelica Buendía**, (Metropolitan Autonomous University), **Alejandra Falabella** (Universidad Alberto Hurtado, Chile), **Carmuca Gómez-Bueno** (Universidad de Granada), **Veronica Gottau** (Universidad Torcuato Di Tella), **Carolina Guzmán-Valenzuela** (Universidad de Chile), **Antonia Lozano-Díaz** (University of Almería), **Antonio Luzon**, (Universidad de Granada), **María Teresa Martín Palomo** (University of Almería), **María Fernández Mellizo-Soto** (Universidad 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), **María Veronica Santelices** (Pontificia Universidad Católica de Chile)

Claudio Almonacid

Universidad Metropolitana de Ciencias de la Educación, Chile

Miguel Ángel Arias Ortega

Universidad Autónoma de la Ciudad de México

Xavier Besalú Costa

Universitat de Girona, España

Xavier Bonal Sarro

Universidad Autónoma de Barcelona, España

Antonio Bolívar Boitia

Universidad de Granada, España

José Joaquín Brunner

Universidad Diego Portales, Chile

Damián Canales Sánchez

Instituto Nacional para la Evaluación de la Educación, México

Gabriela de la Cruz Flores

Universidad Nacional Autónoma de México

Marco Antonio Delgado Fuentes

Universidad Iberoamericana, México

Inés Dussel, DIE-CINVESTAV,

México

Pedro Flores Crespo

Universidad Iberoamericana, México

Ana María García de Fanelli

Centro de Estudios de Estado y Sociedad (CEDES) CONICET, Argentina

Juan Carlos González Faraco

Universidad de Huelva, España

María Clemente Linuesa

Universidad de Salamanca, España

Jaume Martínez Bonafé

Universitat de València, España

Alejandro Márquez Jiménez

Instituto de Investigaciones sobre la Universidad y la Educación, UNAM, México

María Guadalupe Olivier Tellez,

Universidad Pedagógica Nacional, México

Miguel Pereyra

Universidad de Granada, España

Mónica Pini

Universidad Nacional de San Martín, Argentina

Omar Orlando Pulido Chaves

Instituto para la Investigación Educativa y el Desarrollo Pedagógico (IDEP)

José Ignacio Rivas Flores

Universidad de Málaga, España

Miriam Rodríguez Vargas

Universidad Autónoma de Tamaulipas, México

José Gregorio Rodríguez

Universidad Nacional de Colombia, Colombia

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

José Luis San Fabián Maroto

Universidad de Oviedo, España

Jurjo Torres Santomé,

Universidad de la Coruña, España

Yengny Marisol Silva Laya

Universidad Iberoamericana, México

Ernesto Treviño Ronzón

Universidad Veracruzana, México

Ernesto Treviño Villarreal

Universidad Diego Portales Santiago, Chile

Antoni Verger Planells

Universidad Autónoma de Barcelona, España

Catalina Wainerman

Universidad de San Andrés, Argentina

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/FGV), **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

Lílian 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