

Evidence-Based Design Principles for Effective Professional Development: A Critical Appraisal of the Evidence

Christa S. C. Asterhan, Hebrew University Jerusalem, Israel, asterhan@mail.huji.ac.il
Adam Lefstein, Ben Gurion University of the Negev, Israel, lefstein@bgu.ac.il

Abstract: Scholarly efforts to identify core design features for effective professional development (PD) efforts have rapidly grown in the last two decades. Based on extensive literature reviews, meta analyses and large-scale quantitative studies, scholars have arrived at short lists of core design principles for effective PD programs. These design principles are presented as based on strong evidence from large-scale, replicated and rigorous research studies, and as at the heart of consensus among PD scholars. In the present essay, we appraise the quality of the evidence on which this claim is based. We identify several major flaws in the research base on which such claims are based and conclude that, overall, the evidence is weak and claims about strong evidence-based findings is misleading. Additional reservations about this research program are discussed.

In search of evidence-based design principles for professional development

Since the turn of the century, academic and practitioner interest in the systematic study of the effectiveness of formal professional development (PD) efforts has grown rapidly. This surge of interest can be attributed to a mixture of theoretical developments (in theories of learning, interest in life-long learning); budgetary concerns (significant amounts of capital and effort invested in PD, alongside reports about limited returns on these investments, e.g., Harris & Sass, 2011; TNTP, 2015); policy developments (both educational and research funding policies, e.g., Lynch et al., 2019); and growing dissatisfaction with existing PD structures and programs.

Scholars in this field not only ask *whether* formal PD is effective, but also, and even predominantly, attempt to identify *which features* of PD produce positive results (i.e., what works?). The rationale behind this search for effective PD features is straightforward: Once the pivotal features are recognized and defined, these can allegedly be transformed into a set of research-based design principles to guide effective PD policy and program development. In its aims, this rather new field of scholarly interest is similar to a more longstanding research field, often referred to as *Instructional Design* or *Instructional Science* and the like. Instead of focusing on teacher learning, research efforts in this field focus on student learning activities (how to design learning materials, activities and settings for students to optimize learning outcomes). Yet, the ultimate aim is identical, that is: to extract a set of design principles for optimal learning activities. Even though significant knowledge has accumulated about (designing for) student learning, this body of knowledge cannot be simply transferred to teacher learning, at least not without careful, empirical checks. Kennedy (2016) summarizes the current state of the field thus: "Education research is at a stage in which we have strong theories of *student* learning, but we do not have well-developed ideas about *teacher* learning, nor about how to help teachers incorporate new ideas into their ongoing systems of practice." (p. 971).

A second characteristic of this literature is the aspiration to base these design recommendations on the strongest empirical evidence possible, prioritizing research methods that allow causal inference, that are based on large data sets, that use quantifiable, objective student outcomes as the main measure of success, and that have been tested across many empirical studies. Since the ultimate goal of most PD efforts is to improve student learning outcomes, it is not surprising that student test scores have become the "gold standard" for effectiveness studies. However, student outcomes are also a very ambitious measure as they are the most distal variable in a long causal chain of effects: The PD program features are expected to affect teachers' skills, beliefs and/or knowledge during PD, which is expected to translate into differences in participants' classroom practices, which in turn are expected to affect (cognitive, motivational, affective) aspects of student action, which, eventually, are expected to translate into individual student score improvement on objective tests. Some have then also considered more proximal, quantifiable outcome measures, such as changes in observable teaching behavior. Others have argued that, given the long causal chain of the intended effects, even small effects on student achievement should be considered noteworthy (Kennedy, 2016). Moreover, change is expected to take time and to progress incrementally, as teachers adopt and try out new ideas in their practice, and adapt and refine them over time.

Third, a quick reading of the most highly cited publications in this literature shows that there is considerable scholarly agreement about the existence of a core set of features that maximize PD program effectiveness, and that the evidence base for this set is robust, according to the aforementioned criteria of strong and replicated evidence. The identification of a list of features of effective professional development is a recurrent

theme in the professional educational and scientific literature over the course of the past 25 years at least. Guskey (2003) compared 13 such lists published between 1995-2001, noting the divergence between the different lists, despite some of the authors claiming that there now exists "an almost unprecedented consensus... among researchers, professional development specialists, and key policymakers" about the characteristics of effective PD (Hawley & Valli, 1996). More recently, Desimone (2009), synthesized the findings from the available literature at the time by concluding that "there is enough empirical evidence to suggest that there is in fact a consensus on a core set of features"¹ (p. 183):

- 1) a focus on subject matter content and how students learn that content;
- 2) a collective and collaborative effort with colleagues that allows interaction;
- 3) engagement in active learning tasks (such as curriculum design, modeling, enactment, and reflection) as opposed to passive lecturing;
- 4) coherence with existing curricula and policies, as well as teacher beliefs and values;
- 5) sufficient contact hours with participants (Desimone sets the minimum at 20 contact hours, others have argued that longer durations are necessary).

Many other systematic literature reviews and large scale survey studies have arrived at similar conclusions (e.g., Darling-Hammond et al, 2017; Garet et al., 2001; Kennedy, 1998; Penuel, Fishman, Yamaguchi, & Gallagher, 2007; Sher & O'Reilly, 2009; Timperley et al., 2007) and may have even expanded the list with one or two additional features. For example, Darling-Hammond and colleagues (2017) added the use of novel, non-traditional professional development structures (e.g., teacher study groups, coaching) and the availability of expert guidance (e.g., coaches, models). These lists of features have been adopted and reiterated in numerous scholarly publications, reports and PD programs, both as a rationale for designing new PD efforts (Kennedy, 2017), as well as a starting point for subsequent research (e.g., Gibbon & Cobb, 2017). Indeed, we have increasingly found it a rarity when such core features are *not* mentioned as an evidence-based starting point in the introductory section of a report, thesis, dissertation or research publication related to PD. For example, to date (Nov, 2019) the Desimone (2009) paper has been cited 3,400 times according to Google Scholar, Garet et al. (2001) over 6,500 times, and the Darling-Hammond et al (2017) report 422 times in less than two years' time. The numerous citations and reiterations of these effective PD features have further cemented this sense of consensus in the field.

This particular set of features does in fact corroborate with conventional wisdom. Some of them are theoretically well-aligned with findings from longstanding research on (student) learning, namely learning activities that require *active, constructive and interactive* processes; a *constructive and collaborative atmosphere*; and situating learning opportunities in *authentic settings*. Other features from this set appeal to common sense, such as *expert guidance, extensive contact hours, and coherence with existing policies and curricula*. *Focus on content* could also be expected to improve outcomes, especially when increased content knowledge is the focus of teacher and student assessment and in fields in which teachers' content knowledge proficiency is often viewed as in need of improvement (such as elementary science and mathematics, Kennedy, 2016). The addition of "a focus on how *students* learn specific content" to the content focus feature, is particular to the field of teacher PD. It is not included in every list, but a recurrent theme in other strands of educational literature, where teacher noticing and understanding of student thinking has been identified as a pivotal element of professional vision and effective teaching (e.g., Sherin, Jacobs & Philipp, 2011). Moreover, at least one recent study on teacher PD effectiveness has shown that teacher learning of content has a more substantive effect on student achievement when it is integrated into analyses of their everyday teaching practice (Taylor et al., 2017).

However, in spite of this often cited and celebrated consensus, some scholars have critiqued the empirical research on which it is based (Hill et al., 2013; Wayne et al., 2008) or have failed to find such distinctive features when using stricter criteria of inclusion in their reviews (e.g., Kennedy, 2017; Yoon et al., 2007). Moreover, recent studies in which PD has been purposefully designed according to this list of features and then compared to other conditions have not produced the anticipated results (e.g., Garet et al., 2008; 2011; 2016; Penuel et al., 2011).

In the present essay, we aim to explore this apparent discrepancy by delving deeper into the research base from which these sets of core PD features have been derived.

The research base behind the evidence-based, effective PD features

To date, most of the highly cited efforts that sought to identify PD program features broadly fall into one of two categories: reviews of (quasi-)experimental comparisons or large scale, correlational studies of existing variance in PD and outcomes. We discuss them separately.

Reviews of (quasi-) experimental comparisons

The first category contains systematic literature reviews and meta analyses of existing empirical research (e.g., Blank & de las Alas, 2009; Desimone, 2009; Darling-Hammond et al 2017; Gersten et al., 2014; Hawley & Valli, 1999; Kennedy, 1998; Kraft et al, in press; Lynch et al., 2019; Maandag et al., 2017; Sher & O'Reilly, 2009; Timperley et al., 2007; Yoon et al., 2007). Selection criteria require that only (quasi-) experimental studies are considered that contain some form of systematic comparisons between quantified classroom outcomes (teacher behavior and/or student achievement). The overall goal is to stipulate whether PD has an impact and to characterize what sets successful PD programs apart. On first impression, this seems like a commendable endeavor and the fact that different research reviews converge on the same set of PD features is encouraging. However, we identify several serious methodological shortcomings in these reviews:

Some of these shortcomings are almost self-evident. For example, some of the more influential literature reviews *purposefully* choose to consider only studies that reported that reported a positive link between teacher professional development, teaching practices, and student outcomes. For example, Darling-Hammond et al (2017) compiled a set of 35 studies that way, and then identified the most common features of the PD approaches in these studies (e.g., all 35 contained elements of active learning) to arrive at a set of 7 common elements. The seven "effective" elements may also be common to studies that reported null or negative results (e.g., Garet et al., 2008), but these studies were *by design* disregarded.

Others shortcomings are not specific to the execution of a particular review or meta-analysis, but are rather due to the type of research that is available and can be reviewed to answer questions regarding PD effectiveness. Conducting a systematic review on what works in PD features is challenging, because the educational research community has not, until recently, demonstrated much interest in the design of PD as an object of rigorous research in and of itself (Lynch et al., 2019). Yet, much educational research on innovative learning and teaching includes an aspect of PD. In most cases, a team of researchers, or other stakeholders, is interested in exploring the impact of a new curriculum, novel teaching practice, or a set of instructional materials in order to test their effects on teacher classroom behavior or student learning. PD is provided to the participating teachers to assist them in implementing these changes in their classroom teaching. Even though the PD program will likely be designed according to the most up-to-date views on teacher PD and learning, the design in and of itself is not the object of investigation. Thus, in many cases, teachers in the treatment group will have both received some form of PD and also new materials or curricula to improve instruction, whereas teachers in the control condition (if included) received neither PD nor the new instructional practices and are teaching as usual. The PD is not an object of study in itself but a means of achieving change in classroom teaching in order to improve student achievement.

To isolate the benefits of PD in and by itself – i.e., to determine whether PD has an effect – researchers need to ensure that participation in PD is the only difference between groups. Even though the existence of (quasi-) experimental comparisons is a standard requirement for inclusion in the reviews and meta analyses, this is often not further specified to mean that PD needs to be isolated from other artifacts, such as curricular reform. Only few reports explicitly refer to this distinction (e.g., Lynch et al., 2019) or include it as a selection criterion. For example, Yoon et al. (2007) reviewed more than 1,300 studies that were identified as potentially addressing the effect of teacher professional development on student achievement in three key content areas (math, science and language arts). Only nine were found to meet the What Works Clearinghouse evidence standards, included student achievement outcome scores and included a control comparison in which teachers received no PD but could otherwise be considered equal. Taken together, those studies show that PD has a moderate effect on teacher outcomes, specifically on those measures that corroborate with the content of the PD program, and no-to-small effects on student achievement (see also reports from Kennedy, 2016; Garet et al., 2008, 2016, for more recent studies).

However, the question remains whether it is possible to distill *features of effective PD design* from this research base. We argue that this is not possible. To be able to state that a particular PD *feature* is more effective than another, it is not sufficient that PD was isolated from other instruction improvement aspects. This requires that research designs enable comparisons between two (or more) PD programs that differ in terms of program features (e.g., with or without video cases of classroom practice), but are otherwise identical with regard to content and duration (if PD duration is not a variable of interest). We would like to illustrate the importance of this common-sense, yet often overlooked caveat with an illustration: Imagine a group of researchers interested in exploring the effectiveness of project-based learning methods and to this end design a study with two conditions: In the first group, 5th graders learned about photosynthesis through project-based learning, whereas in the second group photosynthesis was not taught at all. Not surprisingly, students in the first group showed better understanding of photosynthesis than children in the control group. Even if this study were replicated a hundred times in 20 different countries with thousands of students in each trial, we would all still agree that this research

design does not enable us to conclude that project-based teaching is superior to other teaching methods. At best, we can infer that teaching photosynthesis in this way is better than not teaching it at all.

To the best of our knowledge, and perhaps not surprisingly given the recency of empirical interest in effective PD activity features and the enormous amount of effort and skill that are required to conduct it, studies in which the effects of different design features of PD on teacher or student outcomes are compared directly are very rare. Therefore, most literature reviews and meta-analyses, even the more rigorous or recent ones, rely on research that was not intended to, and whose design cannot, provide solid answers to questions concerning core features of effective PD design. Some of the reports we have read rely solely on this type of research (e.g., Gersten et al., 2014). But, even more recent reviews and reports that (inadvertently?) included several studies with direct PD design comparisons failed to distinguish between them in any way or give them any special status, and in any case the majority of the research included was of the PD vs. no-PD comparison type. The conclusions on which these reviews and meta analyses are based are then analogous to the faulty conclusion from the imaginary illustration on the superiority of project-based learning described above.

In addition to these serious flaws in existing reviews and meta analyses, findings from recent studies that did include direct and controlled comparisons between different formats of PD delivery on the same topic in fact failed to find differences (Heller et al, 2012; Fisher, Schumaker, Culbertson, & Deshler, 2010; Fishman et al., 2013; Garet et al., 2011; Grigg, Kelly, Gamoran & Borman, 2013; Osborne, Borko, Fishman et al, 2019; Powell et al., 2010; Russell, Carey, Kleiman, & Venable, 2009; but see Taylor et al., 2017 for an exception)². Some of these were even designed specifically to test the effectiveness of specific PD program features identified in the literature (e.g., Garet et al., 2008). Taken together, the different PD program design features that were targeted in these studies were not found to affect the outcomes of PD differently, particularly not *student* outcomes. For example, Garet et al (2008) compared a standard PD program for early reading interventions with two PD programs specifically designed to integrate the effective PD recommendation of “content-focus” (both), as well as continuous one-on-one coaching sessions (in only one program). The three PD programs produced similar outcomes. Osborne et al (2019) compared three versions of a PD program for improving argumentation in science classes. The three versions improved teacher facilitation of classroom argumentation equally well.

Large scale correlational studies on reform initiatives

The second category of scholarly work that has aimed to identify features of effective PD is characterized by a different research approach. They are set in large-scale reform efforts that allow for local decisions about the exact form of PD delivery and implementation. In these cases, topical content can be assumed to be held fairly constant (provided and stipulated by educational administrators), but delivery format of PD is not. Data on this existing variability is then collected and correlated with student or teacher outcome measures (e.g., Cohen & Hill, 2001; Desimone et al., 2002; Fischer et al, 2018; Garet et al., 2001; Penuel et al., 2007). However, a major drawback of this group of studies is that, notwithstanding the often awe-inspiring data collection efforts, the complex statistical modeling, and rigorous analyses, in the end these are survey-based studies that rely almost entirely on self-report data: Participating teachers are asked to report on both the content of their professional development experiences, as well as the key outcomes of the program, such as perceived knowledge increases, perceived improvements in teaching capacity, or perceived changes in instructional practices (e.g., What was the impact of the PD on your classroom teaching practices?). For example, in the widely cited Garet et al. (2001) study, teachers’ self-reported increases in knowledge and skills from PD were correlated with their reports on features of the PD (e.g., extent of focus on content, coherence with goals and expectations, and duration).

Conclusion

Scholarly interest in the effectiveness of (different forms of) teacher PD as a distinct topic for empirical research—as opposed to using PD as a means to study other topics—is relatively new. Based on extensive literature reviews and large-scale quantitative studies from the last two decades, a short list of core design principles for effective PD programs have been identified. These are often times presented as based on strong evidence from large-scale, replicated and rigorous research studies, and as at the heart of consensus among PD scholars. These principles corroborate with common sense, existing views and trends on meaningful PD, as well as socio-constructivist, situative and cognitive theories of learning. They also align with findings from other fields of research on teacher PD, such as the literature on professional learning communities and teacher on-the-job learning (e.g., Bryk & Camburn, & Louis, 1999; Louis & Marks, 1998).

Nevertheless, in spite of the impression created in many and often highly cited publications, we conclude that the empirical basis that has been used to arrive at these effective PD design features is weak. The impression that has been created in the PD literature that this list of features is backed by a strong evidence base that adheres to the strictest criteria of research quality is misleading. Not because the studies on which the extensive reviews

are based were weak or badly executed studies *per se*, but because most of them were never designed to test the question of PD effectiveness. It would be more accurate to state that an initial set of features has been identified, which needs to be further explored, tested and specified in research that explicitly compares different sets of PD features (see also Kennedy, 2016).

With the recent interest in teacher learning as a topic of empirical interest in and by itself, such controlled studies are gradually becoming more frequent (e.g., Osborne et al., 2019; Taylor et al., 2017). Thus far, however, findings from these studies do not echo the expectations about differential effects for different PD design features: Even though PD often had an overall effect on student and/or teacher outcomes, different PD program designs on the same topic produced roughly equal outcomes. Even though it is still early days and null results cannot prove that the design features do not matter at all, this lack of effects from direct and controlled comparisons do raise further questions concerning the effectiveness of popular PD design features.

We have critiqued here the methodological shortcomings of the empirical foundation on which claims about the identification of evidence-based, effective PD design features have been based. In order to make such claims, more comparative research between different PD designs is needed. However, in closing, we wish to also problematize some of the assumptions underlying the entire program that seeks to uncover effective PD design principles and features, in order to design PD accordingly:

First, it is unrealistic to expect to find a one-size-fits-all answer to the question of what makes PD effective. Effectiveness is, among other factors, a product of the interaction of means and ends (different design features for different types of knowledge and skills). For example, PD focusing on teaching aspect X may be more effective through direct instruction and individual exercises to improve fluency, whereas changing aspect Y can best be accomplished through intensive hands-on simulations or collaborative peer reflection.

Second, effectiveness also interacts with what aspect of teaching the PD aims to address. Some aspects are easier to change than others and some instructional reforms are more likely to impact student achievement than others. For example, it may prove to be much easier to train teachers in the use of classroom management routines, which can be straightforwardly exemplified and practiced, than the facilitation of classroom discussions, which requires greater knowledge, flexibility, and judgement. This issue further confounds the comparison of findings regarding core design features.

Third, the role of the facilitator, of expert and experienced PD leaders and developers is likely to be crucial (see also Kennedy, 2016), but it is much more difficult to characterize or measure the different forms of expertise (relative, to characterizing, for example, PD program focus, activities or duration).

Fourth, this essay focused on concerted efforts to improve teaching through professional development programs, which are usually initiated, organized and conducted by education reformers or researchers external to the schools in which teachers work. However, teachers also learn on the job, through natural processes of engaging in their work and accumulating experience. Indeed, since teaching involves a great deal of tacit knowledge and situational judgement, much of our development as teachers involves learning from experience (Eraut, 1994; Shulman, 2004). A comprehensive design for teacher PD would be well-advised to integrate structured PD programs with ongoing teacher workplace learning. So, for example, rather than treating a 30 or 60 *hr* workshop on some curriculum or teaching strategy as a stand-alone program, we would likely benefit from ways of integrating such workshops with programs to support teachers' ongoing trial of program ideas and reflection on their progress within their school-based professional communities.

Finally, and perhaps most importantly, merely trying to identify a set of features will not be enough. Effectiveness depends on how these features are enacted (see Kennedy 2016). This is well known in the more mature field of Instructional Science. Take for example student peer group work. Yes, group work can be effective for certain types of student learning outcomes, but its effectiveness is dependent, among other conditions, on the type of task that students are assigned, students' collaboration and communication skills, and the availability and quality of teacher facilitation. Likewise, with teacher professional development: research on effectiveness can usefully inform this work, but ultimately the professional knowledge, skill, judgement, and wisdom of practitioners and participants will critically shape its outcomes.

Endnotes

- (1) Interestingly, seven years prior, Desimone and colleagues (2002) presented the evidence base for this consensus in a much different light. After noting that a "professional consensus is emerging about particular characteristics of "high quality" professional development" and detailing these characteristics, they express the following reservation: "Although lists of characteristics such as these commonly appear in the literature on effective professional development, there is little direct evidence on the extent to which these characteristics are related to better teaching and increased student achievement."

- (2) A few studies in which different forms and duration of PD were compared were left out of this list as they contained study-specific confounds or other specific issues that rendered the interpretations of their results equivocal or problematic (e.g., Landry et al., 2009; Kleickman et al., 2016; Roth et al., 2011).

References

- Blank, R. K., & de las Alas, N. (2009). The effects of teacher professional development on gains in student achievement: How meta-analysis provides scientific evidence useful to education leaders. Washington D.C.: Council of Chief State School Officers.
- Bryk, A. S., Camburn, E., & Louis, K. S. (1999). Professional community in Chicago elementary schools: Facilitating factors and organizational consequences. *Educational Administration Quarterly*, 35(5), 751–781.
- Darling-Hammond, L., Hyler, M. E., Gardner, M. (2017). *Effective teacher professional development*. Palo Alto, CA: Learning Policy Institute.
- Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, 38(3), 181-199.
- Desimone, L. M., Porter, A.C., & Garet, M.S., Yoon, K.S., & Birman, B.F. (2002). Effects of professional development on teachers' instruction: Results from a three-year longitudinal study. *Educational Evaluation and Policy Analysis*, 24(2), 81-112.
- Fischer, C., Fishman, B., Dede, C., Eisenkraft, A., Frumin, K., Foster, B., ... & McCoy, A. (2018). Investigating relationships between school context, teacher professional development, teaching practices, and student achievement in response to a nationwide science reform. *Teaching and Teacher Education*, 72, 107-121
- Fisher, J. B., Schumaker, J. B., Culbertson, J., & Deshler, D. D. (2010). Effects of a computerized professional development program on teacher and student outcomes. *Journal of Teacher Education*, 61, 301–312.
- Fishman, B., Konstantopoulos, S., Kubitskey, B. W., Vath, R., Park, G., Johnson, H., & Edelson, D. C. (2013). Comparing the impact of online and face-to-face professional development in the context of curriculum implementation. *Journal of Teacher Education*, 64(5), 426–438.
- Garet, M. S., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W., ... & Zhu, P. (2008). The impact of two professional development interventions on early reading instruction and achievement. NCEE 2008-4030. National Center for Education Evaluation and Regional Assistance ; Institute of Education Sciences, U.S. Department of Education.
- Garet, M. S., Heppen, J. B., Walters, K., Parkinson, J., Smith, T. M., Song, M., ... & Borman, G. D. (2016). Focusing on Mathematical Knowledge: The impact of content-intensive teacher professional development. NCEE 2016-4010. National Center for Education Evaluation and Regional Assistance; Institute of Education Sciences, U.S. Department of Education.
- Garet, M.S., Porter, A.C., Desimone, L., Birman, B.F., & Yoon, K.S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, 38(4), 915-945.
- Garet, M. S., Wayne, A., Stancavage, F., Taylor, J., Eaton, M., Walters, K., . . . and Doolittle, F. (2011). Middle school mathematics professional development impact study findings after the second year of implementation. (No. NCEE 2011-4024). Washinton, DC.: IES; National Center for Education Evaluation and Regional Assistance; Institute of Education Sciences, U.S. Department of Education.
- Gersten, R., Taylor, M. J., Keys, T. D., Rolffhus, E., & Newman-Gonchar, R. (2014). Summary of research on the effectiveness of math professional development approaches. (No. REL 2014-74010).Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast.
- Gibbons, L. K., & Cobb, P. (2017). Focusing on teacher learning opportunities to identify potentially productive coaching activities. *Journal of Teacher Education*, 68(4), 411-425.
- Grigg, J., Kelley, K. A., Gamoran, A., & Borman, G. D. (2013). Effects of the scientific inquiry professional development interventions on teaching practice. *Educational Evaluation and Policy Analysis*, 35(1), 38-56.
- Guskey, T. R. (2003). Analyzing lists of the characteristics of effective professional development to promote visionary leadership. *NASSP bulletin*, 87(637), 4-20.
- Harris, D.N. & Sass, T.R. (2011). Teacher Training, Teacher Quality and Student Achievement. *Journal of Public Economics* 95 (7–8): 798–812.
- Hawley, W. D., & Valli, L. (1996). The essentials of effective professional development: A new consensus. *Professional Development Newsletter–ASCD Human Resource Development Program*, 1-2.

- Heller, J. I., Daehler, K. R., Wong, N., Shinohara, M., & Miratrix, L. W. (2012). Differential effects of three professional development models on teacher knowledge and student achievement in elementary science. *Journal of Research in Science Teaching*, 49(3), 333–362
- Hill, H.C., Beisiegel, M., & Jacob, R. (2013). Professional development research: Consensus, crossroads, and challenges. *Educational Researcher*, 42(9), 476-487.
- Jacobs, V., Franke, M., Carpenter, T., Levi, L., & Battey, D. (2007). Professional development focused on children's algebraic reasoning in elementary school. *Journal for Research in Mathematics Education*, 38(3), 258–288.
- Kennedy, M.M. (1999). The role of preservice teacher education. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 54-85). San Francisco, CA: Jossey Bass.
- Kennedy, M. M. (2016). How does professional development improve teaching?. *Review of Educational Research*, 86(4), 945-980.
- Kraft, M. A., & Papay, J. P. (2014). Can professional environments in schools promote teacher development? Explaining heterogeneity in returns to teaching experience. *Educational Evaluation and Policy Analysis*, 36(4), 476-500.
- Louis, K. S., & Marks, H. M. (1998). Does professional community affect the classroom? Teachers' work and student experiences in restructuring schools. *American Journal of Education*, 106(4), 532-575.
- Lynch, K., Hill, H. C., Gonzalez, K. E., & Pollard, C. (2019). Strengthening the Research Base that Informs STEM Instructional Improvement Efforts: A Meta-Analysis. *Educational Evaluation and Policy Analysis*, 41(3), 260-293.
- Maandag, D. W., Helms-Lorenz, M., Lugthart, E., Verkade, A. T., & Van Veen, K. (2017). Features of effective professional development interventions in different stages of teacher's careers. A review of empirical evidence and underlying theory.
- Osborne, J. F., Borko, H., Fishman, E., Gomez Zaccarelli, F., Berson, E., Busch, K. C., ... & Tseng, A. (in press). Impacts of a Practice-Based Professional Development Program on Elementary Teachers' Facilitation of and Student Engagement With Scientific Argumentation. *American Educational Research Journal*.
- Penuel, W. R., Fishman, B. J., Yamaguchi, R., & Gallagher, L. P. (2007). What makes professional development effective? Strategies that foster curriculum implementation. *American Educational Research Journal*, 44(4), 921-958.
- Penuel, W. R., Gallagher, L. P., & Moorthy, S. (2011). Preparing teachers to design sequences of instruction in earth systems science: A comparison of three professional development programs. *American Educational Research Journal*, 48(4), 996-1025.
- Powell, D. R., Diamond, K. E., Burchinal, M. R., & Koehler, M. J. (2010). Effects of an early literacy professional development intervention on head start teachers and children. *Journal of Educational Psychology*, 102, 299
- Russell, M., Carey, R., Kleiman, G., & Venable, J. D. (2009). Face-to-face and online professional development for mathematics teachers: A comparative study. *Journal of asynchronous learning networks*, 13(2), 71-87.
- Taylor, J. A., Roth, K., Wilson, C. D., Stuhlsatz, M. A., & Tipton, E. (2017). The effect of an analysis-of-practice, videocase-based, teacher professional development program on elementary students' science achievement. *Journal of Research on Educational Effectiveness*, 10(2), 241-271.
- Timperley, H., Wilson, A., Barrar, H., & Fung, I. (2007). *Teacher professional learning and development. Best evidence synthesis iteration (BES)*. Wellington, New Zealand: Ministry of Education.
- TNTP (2015). *The mirage: Confronting the hard truth about our quest for teacher development*. Brooklyn, NY: TNTP.
- Wayne, A. J., Yoon, K. S., Zhu, P., Cronen, S., & Garet, M. S. (2008). Experimenting with teacher professional development: Motives and methods. *Educational Researcher*, 37(8), 469-479.
- Yoon, K.S, Duncan, T., Lee, S.W.Y., Scarloss, B., & Shapley, K.L. (2007). *Reviewing the evidence on how teacher professional development affects student achievement*. Washington, DC: Institute for Education Sciences.