

# Teacher as Researcher

12.19.2019

Robert J. Marzano, Danette Parsley, Douglas J. Gagnon, and Jennifer S. Norford

For decades, the notion of teachers engaging in research has been discussed and carried out under the heuristics and methodologies of action research (Manfra, 2019; Pine, 2009). A typical action research project might involve an individual teacher studying the effectiveness of a specific instructional strategy like having students preview content before receiving direct instruction. Although teachers are frequently encouraged to engage in such projects, action research is seldom considered a legitimate form of research. This disregard for action research is mainly because the findings of these projects do not generalize beyond the context of the teacher and classroom in which they were conducted. The position of the research community has been that situated findings are not very useful.

Experiments provide the best evidence with respect to treatment effects; they can, however, yield results that are local and particular. Most researchers, however, are interested in knowing whether these effects generalize to other populations and settings. They may also want to know whether such effects generalize to other outcomes and treatment implementations. Researchers often rely on a combination of approaches to maximize the generalizability of their results... Statistically, the only formal basis for ensuring the generalization of causal effects is to sample from a well-defined population (Schneider et al., 2007, pp. 28-29).

This perspective is necessary and valid for experiments designed to be highly generalizable, but teacher-conducted experiments are designed to be generalizable only within the local context of the individual teacher carrying out the experiment. From this focused perspective, experiments designed and conducted by individual teachers for the purpose of studying the relationship between their actions in the classroom and specific academic or non-academic outcomes for their students can be considered a viable form of research. In fact, such focused studies might contribute as much, if not more, to the advancement of a scientific approach to classroom instruction than experiments designed for maximum generalizability. Our central thesis is that individual classroom teachers can be legitimate researchers when it comes to conducting experiments on the effectiveness of specific instructional strategies. This notion calls for a reconceptualization of many of the elements associated with experiments.

## Instructional Strategies as Unique Interventions

Instructional strategies are a unique type of intervention because their effects are observable in a relatively short period of time. As such, they are perfect candidates for teacher experiments. Consider the instructional strategy of previewing, which involves activities that provide students with an overview of the information they are about to receive via some form of instruction. The intended outcome of previewing is that students better comprehend new information provided to them and the effect of this strategy should be observable within a single lesson or a few lessons. Such a *short-cycle* outcome is substantially different from an outcome found in an experiment designed to examine the effect of a year-long intervention such as a new writing program. Because instructional strategies often have a short-cycle effect on student learning, they are poor candidates for experiments that last for an entire year or even an entire semester or quarter.

The effective use of previewing strategies before students read a textbook chapter will be observable when students are assessed on their understanding of the content in that chapter. However, that concrete and observable effect will not be apparent when students are administered a test at the end of

the quarter, the end of the semester, or the end of the year. The classroom teacher is in an ideal position to discern and measure such short cycle effects.

## Examining Causality as It Relates to Instructional Strategies

Teacher-designed experiments have the same basic purpose as experiments designed to foster generalizations that can be applied across a broad population: to identify causal relationships. Holland (1986) notes that although experiments are not the only type of study to disclose causal relationships, they are the simplest setting in which to do so.

The quest for causality is foundational to experimentation. In 1690, John Locke published *An Essay Concerning Human Understanding*. There, he attempted to describe the basic dynamics of human learning and human understanding, both of which involve identifying causes and effects which Locke defined in the following way: “A *cause* is that which makes any other thing, either simple idea, substance, or mode, begin to be; and an *effect* is that which had its beginning from some other thing” (1690/1975, p.325).

At a fundamental level, teacher-designed experiments are efforts to establish a relationship between something a teacher does (the cause) and a specific student outcome (the effect). Since Locke’s treatise in 1690, researchers have been devising ways to model and measure causal relationships. For example, Rubin (1974, 1977, 1978) defined a causal effect as the difference between what would happen to a participant (i.e., a human subject) in the experiment under the treatment condition and what would happen to that participant if he or she had instead been exposed to the control condition. The difference between a control condition and a treatment condition can be defined mathematically in the following way:

$$\Delta u = Y_{tu} - Y_{cu} \text{ (Schneider et al., 2007).}$$

Here,  $Y_{tu}$  is the score on the measured outcome for participants ( $u$ ) of the study after they have experienced the condition involving the treatment ( $t$ );  $Y_{cu}$  is the score for participants ( $u$ ) without the treatment condition. The without treatment condition is referred to as the control condition and signified as  $c$ . This model is referred to as the counterfactual account of causality. About the model, Schneider and colleagues explain:

While this definition provides a clear theoretical formulation of what a causal effect is, it cannot be tested empirically because if we have observed  $Y_{tu}$  we cannot also observe  $Y_{cu}$ . This is often referred to as the fundamental problem of causal inference (p.13).

Holland (1986) explains that there are two general solutions to solving the fundamental problem of causal inference: the scientific solution and the statistical solution.

Researchers use two approaches to implement the scientific solution. The first approach is to observe the participant ( $u$ ) at two points in time: (1) before the participant has experienced the experimental treatment, and (2) after the participant has experienced the experimental condition. The causal effect in this case is the difference between the outcome for participant  $u$  after the treatment condition and the outcome the participant displayed under the control condition.

There are important assumptions that must be made within this approach. One is *temporal stability* which means that there is consistency of response by the participant across time. In the absence of a treatment, the participant’s response is always the same. Another is *causal transience*, which means that the effect of the treatment condition ( $t$ ) does not affect the participant’s response to the control condition. These assumptions make little sense with human participants in the social sciences. Because

we expect that humans may provide inconsistent responses over time, due to any number of factors (e.g., mood, environment), and we expect a treatment may affect how humans respond to the control condition, researchers cannot use this method within the scientific approach with human participants.

The second approach within the scientific solution is to assume that all participants in the experiment are identical in all respects. It therefore makes no difference which participant receives the treatment condition. This assumption is referred to as *unit homogeneity* and also makes little sense with human subjects—individuals vary across any number of factors such as gender, age, ethnicity, and so on.

Because neither of the two methods of the scientific solution can be used with human subjects, social scientists use the statistical solution to the fundamental problem of causal inference. According to Schneider and colleagues (2007), the statistical solution focuses on the average causal effect across a population of participants as opposed to focusing on a single participant. The mathematical model to this solution is:

$$\Delta = (Y_t - Y_c)$$

Here  $Y_t$  is the average outcome for participants in the treatment group and  $Y_c$  is the average outcome for participants in the control group. The major assumption that must be made for this solution to work is that:

...individuals or organizational elements (e.g., classrooms or schools) in the treatment and control groups should differ only in terms of treatment group assignment, not on any other characteristics or prior experience that might potentially affect their responses (Schneider et al., 2007, p. 15).

Another assumption that must be made for the statistical solution to work is that random assignment of participants to treatment and control conditions ensures that treatment and control groups differ only in terms of treatment group assignment:

However, if students are randomly assigned to treatment and control conditions, one could expect that treatment group assignment would, *on average*, over repeated trials, be independent of any measured or unmeasured pretreatment characteristic. Because random assignment assures, in expectation, equivalence between groups on pretreatment characteristics, if students in the treatment group score higher on post-treatment assessment...the researcher can conclude, at least in large samples, that this effect is due to differences in the program of instruction rather than to differences in characteristics of students in the two groups (Schneider et al., 2007, p. 15).

All these assumptions represent the ideal situation, which is rarely achieved when experiments are conducted in classrooms. Teacher-conducted research (designed to be generalized within the local context) fairs well as a legitimate form of experimentation when one compares its strengths and weaknesses with those of traditional methods for experiments (designed to be generalized across teachers, schools, and districts). Here we discuss two of the most salient factors that influence the outcome of experiments: generalizability and the validity of the criterion measure.

## Generalizability

Generalizability and random assignment are intimately connected:

Statistically, the only formal basis for ensuring generalization of causal effects is to randomly sample from a well-defined population (not to be confused with the random assignment of participants to treatment and control groups). This is accomplished through an enumeration of

the population of interest (e.g., the U.S. population of high school students) (Schneider et al., 2007, p. 29).

Unfortunately, when the population of interest is broad, like all U.S. high school students, it is almost impossible to randomly select from the population, especially when one considers the myriad of factors on which such students will vary. Indeed, systematic variation in achievement outcomes occur at the individual level, school level, regional level, by urbanicity, and students' stage in life. Thus, sampling from a well-defined population that is broad in nature involves many factors and many levels. In their paper, *Intraclass Correlation Values for Planning Group Randomization in Education*, Hedges and Hedberg (2007) noted that the more one considers the variation across the factors associated with a broad population, the more complex the issue of random assignment becomes. Additionally, failing to account for the variance associated with the many factors embedded in a target population of students will severely affect the statistical power of experiments that are designed to estimate the effects of an intervention on a target population.

The sampling of subjects into experiments via statistical clusters introduces special considerations that need to be addressed in the analysis. For example, a sample obtained from  $m$  clusters (such as classroom or schools) of size  $n$  randomized into a treatment group is not a simple sample of  $nm$  individuals, even if it is based on a simple random sample of clusters. Consequently the sampling distribution of statistics based on such clusters is not the same as those based on simple random samples of the same size (Hedges & Hedberg, 2007, p.3)

The individual teacher conducting an experiment does not face this level of sampling complexity. For example, at the elementary level, a teacher might have a single group of students for the entire year. In a secondary school, a teacher might have four to six groups (i.e., classes) of students. Because the classroom teacher likely has all students in the same school, there is no need to account for school-level clustering.

Additionally, to the extent that other classes under the teacher's purview are similar to the class in which the teacher conducted the experiment, the results will have reasonable application to those other classes. In other words, the experiments conducted on these subjects are highly generalizable to those other classes. This generalizability across classes is particularly important for schools that serve populations that may be unique or underrepresented in the research such as highly mobile students, English learners, or rural schools. Indeed, scholars have argued that what counts as evidence in education must be translated and interrogated by educators to account for unique contexts (Eppley, Azano, Brenner & Shannon, 2018).

A useful and interesting perspective regarding the issue of generalizability within experiments conducted by teachers is provoked by the question: "What is the evidence that large-scale experiments are good vehicles to create highly generalizable conclusions about instructional strategies in a classroom?"

Stated differently, the utility of large-scale experiments intended to be generalizable across large populations (e.g., all high school students) may be severely limited when it comes to specific instructional strategies in the classroom. As described above, an instructional strategy has a short cycle effect on students that might be as circumscribed as a single lesson or even a part of a single lesson. Single lessons are situated within a wide variety of factors that include:

- A specific topic
- Within a specific subject area
- At a specific grade level

- Taught by a specific teacher
- At a specific time of year

Hedges and Hedberg (2007) have estimated the intraclass correlations (ICCs) of a variety of more general factors for which one would have to control to maximize the generalizability of findings. They used data from national databases like the Early Childhood Longitudinal Survey, the National Educational Longitudinal Study of the Eighth Grade Class of 1988, and the Longitudinal Study of American Youth. Their analysis assumed that schools were assigned to treatments. They addressed four dimensions of intervention designs and used the ICC as the metric of study (Table 1).

**Table 1. Dimensions of intervention designs**

Model	Involves
Unconditional model	Testing treatment effects with no covariates. Typically used in settings where the researcher has little opportunity to collect prior information about the individuals participating in the experiment.
Conditional model	Testing of treatment effects conditional on descriptive factors such as gender, race/ethnicity, and socio-economic status. Typically used when the researcher obtains prior data from administrative records.
Residualized unconditional model (or residualized gain model)	Testing treatment effects using pretest scores on the same achievement domain as the outcome measure (e.g., mathematics, reading).
Residualized conditional model	Testing treatment effects using pre-test scores on the same achievement domain as covariates, as well as controlling for descriptive factors such as gender, race/ethnicity and socio-economic status.

Among many trends, Hedges and Hedberg (2007) found that as more lower-level factors (e.g., student-level) were accounted for in a study's design—as found in the second, third, and fourth models—the lower the ICC and the less power the designs had to detect effects. Moreover, given the variety of factors that must be incorporated into short-cycle experiments (e.g., specific topic, time of year), this consequence of reduced power is magnified in the case of short-cycle experiments. Consequently, a model with the necessary controls and sufficient power will require a sample size that is unrealistic to achieve for short-cycle experiments. On the other hand, if all necessary controls are *not* included, it becomes more difficult to trust that effects generalize to a given classroom. Snijders and Bosker (2004) echo the problem of unrealistic sample sizes when they explain that as within group homogeneity increases, the ICC might increase, but the need for more subjects also increases if hierarchical linear modelling is used (p. 24). Of particular importance to this discussion, their analyses did not include many of the control factors that would most likely be of importance to classroom teachers concerned with the validity and utility of a specific instructional strategy. As mentioned above, these include factors such as a specific topic within a domain, teacher experience with a strategy, student engagement in instruction, and so on. In effect, a large-scale study that was focused on specific instructional strategies would need to include a long list of these other factors to produce findings that are meaningful to individual

teachers. The complexity of the design and the sample size required to detect minimally important effect sizes would make it impractical to carry out such a study.

## Validity of the Criterion Measure

Another factor on which teacher-conducted experiments compare well to large-scale experiments is the validity of the criterion measure. According to Raudenbush (2005), this issue might be one of the biggest challenges to the validity of the findings from experimental research, particularly as it relates to instructional interventions.

Indeed, one might argue that a failure to attend systematically to this process of creating good outcome measures is the Achilles heel of evaluation research on instructional innovation. If the process is ignored, trivialized, or mismanaged, we'll be measuring the wrong outcome with high reliability, the right outcome with low reliability, or, in the worst case, we won't know what we are measuring. If we don't know what we are measuring, the causal question (Does the new intervention improve achievement?) is meaningless. If we measure the right outcome unreliably, we will likely find a new program ineffective even if it is effective. If we measure the wrong outcome reliably, we may find that the intervention "works," but we'll never know whether it works to achieve *our* goals (Raudenbush, 2005, p. 29).

Given the short-cycle effect of instructional strategies, it is difficult to find standardized measures of their effectiveness. Most standardized assessments are designed to cover multiple types of knowledge and skill relative to a specific subject area, and focus on more distal outcomes such as student learning. The appropriate criterion measure relative to a given instructional strategy like previewing must be specific to the content and situation for which the teacher engaged students in the strategy. This likely requires using a more proximal outcome than standardized assessments can capture, such as the activation of prior knowledge in the case of previewing. Teacher-conducted research provides for more flexibility and creativity in selecting an outcome measure that is valid in assessing the short-cycle effect of instructional strategies.

## The New Paradigm for Teacher as Researcher

The discussion above implies a new paradigm for experimentation involving instructional strategies at the individual teacher level. As described above, that new paradigm will involve new ways to define sampling procedures, generalizability, and validity for criterion measures. That paradigm must also include new interpretations and new parameters for constructs that include:

- Experimental designs appropriate to individual teachers
- Appropriate statistical procedures
- The relevance of outcomes

### Experimental designs appropriate to individual teachers

The new paradigm for teacher as researcher is likely to include a small subset of the designs that are available in large-scale research efforts. Specifically, the most common design that can be used by teachers would involve an experimental group of students and a control group of students. Preferably, these would come from the same class and be randomly assigned to experimental and control conditions. The short-cycle effect of specific instructional strategies allows for this situation in that many strategies would be expected to produce observable effects in one or two class periods. Those students who were in the experimental condition might first receive instruction with the benefit of the target

strategy. While this treatment was occurring, the students in the control group would work independently in some other room (e.g., the library) on learning activities that enhance their knowledge but do not contaminate their prior knowledge of the content. After the one or two class periods, experimental and control groups would switch and the control group would experience the same one or two class period instruction, this time with the benefit of target instructional strategy.

### Appropriate statistical procedures

The paradigm for teacher-conducted experiments would also be limited in terms of the statistical procedures that are appropriate. For example, in the Instructional Improvement Cycle (Cherasaro, Reale, Haystead, & Marzano, 2015), teachers are required to have a pre-test and post-test for each student. The pre-test is used to test for baseline equivalence. If experimental and control groups do not demonstrate baseline equivalence, teachers are encouraged to go no further in the experiment. Another common (but less rigorous) statistical approach is to use an ANCOVA design in which the pre-test is used as the covariate and the post-test as the criterion measure.

One consideration for the statistical procedures for teacher-designed experiments is the alpha level (i.e., significance level) used to reject the null hypothesis in large-scale experiments. The alpha level is typically set at less than .05 to adequately guard against type I errors. Such a criterion seems reasonable since the conclusions of large-scale experiments are intended to be generalizable across a broad population. In this context, type I errors could have serious consequences because they magnify the consequences of concluding that an intervention “works” when an effect was detected. However, when an experiment is meant to be generalized to the behaviors of an individual teacher, the consequences of making a type I error are less daunting. Moreover, given the smaller sample sizes in teacher-conducted research, type I errors are less likely to occur in the first place.

### The relevance of outcomes

Teacher-designed experiments are much more amenable to focusing on outcomes that are relevant to specific teachers (i.e., those conducting the experiments) than are large-scale experiments on specific instructional strategies. There are multiple reasons for this. For one, teacher-designed experiments are better able to detect the short-cycle effect of specific instructional strategies. Further, such experiments can address specific topics from specific subject areas that are the focus of a small set of lessons. In addition, teachers can include outcomes that directly relate to how they teach or how they might teach. Such outcomes might include student interest, confidence, and level of effort in learning specific topics.

## A National Approach

The new paradigm for teacher as researcher must be constructed in an inductive way by conducting multiple teacher-designed experiments and then examining those experiments to glean generalizations about the nature of the practice and recommendations for conducting effective teacher-designed research.

It is feasible to imagine a national effort that consolidates the findings of multiple teacher-designed experiments. Such a national effort might require the following elements:

- A list of instructional strategies culled from research-based instructional models or frameworks (e.g., Danielson, 2011, 2013; Hattie, 2008; Marzano, 2017)
- Common design specifications (see Instructional Improvement Cycle; Cherasaro, et al., 2015)
- A common data analysis approach (see Instructional Improvement Cycle; Cherasaro, et al., 2015)
- Results sent to common location



- Computation of common effect sizes for the same strategy, then broken down by mediating and moderating variables (e.g., teacher's familiarity with the instructional strategy being studied, length of time strategy was employed)

On a small scale, Robert J. Marzano applied this approach and captured findings in an online database of instructional strategies (Meta-Analysis Database of Instructional Strategies, n.d.) From 2004–11, Marzano collected data from more than 500 teachers through action research at 87 schools in 26 districts. The database includes action research studies conducted by 500 teachers on 22 instructional strategies.

## Building the New Paradigm for Teacher as Researcher

Ideally, the new paradigm will be built in the context of contemporary approaches to teacher collaboration, including professional learning communities (PLCs) and professional learning networks (PLNs).

Since the 1990s, there has been a growing emphasis on creating structures, processes, routines, and expectations for teachers working together to hone their practice and improve outcomes for students, classrooms, and schools. The PLC movement has resulted in a nearly ubiquitous acceptance of teacher collaboration as an essential ingredient for school improvement. As Hargreaves (2018) notes, the “important debates about collaboration now are no longer about whether it is a good thing or not but about how to undertake it with precise designs that promote inquiry, reflection, better practice and increased commitment to change” (p. xxii).

There is no doubt that instructional change and improvement happens at the individual class and school level. However, school systems cannot achieve sustained improvement in teacher practice and student learning outcomes at any scale if it is solely dependent on the professional prowess of individual classroom teachers. Teacher collaboration has been gaining momentum because teachers' participation in quality collaborative learning leads to enhanced human capital (knowledge, skills, attitudes), which helps spread change and improvement throughout the system.

### Professional Learning Communities

PLCs have roots in the literature on professional collaboration (Rosenholtz, 1991) as well as reflective practice (Schön, 1983; Stenhouse, 1975). The term PLC was used in education in the 1990s (Cuban, 1992; Hord, 1997; Louis, Marks, and Kruse, 1996; McLaughlin, 1993), but it became popular in the first decade of the 21st century (Dufour & Eaker, 1998; DuFour, DuFour, and Eaker, 2008; DuFour, DuFour, Eaker and Many, 2010). The purpose of the PLC model is for educators in schools to operate collaboratively in ways that have a direct, positive influence on student learning.

One manifestation of such collaborative endeavors is for teachers to engage in lesson study. In Japan, lesson study is part of *kounaikenshuu*, a comprehensive, school-based approach to professional development that formed the crux of school improvement made popular in the United States in the 1990s (Lewis, 2002).

One of the most common components of *kounaikenshuu* is lesson study (*jugyuu kenkyuu*). In lesson study, groups of teachers meet regularly over long periods of time (ranging from several months to a year) to work on the design, implementation, testing, and improvement of one or several “research lessons” (*kenkyuu jugyuu*). By all indicators, lesson study is extremely popular and highly valued, especially at the elementary level. It is the linchpin of the improvement process (Stigler & Heibert, 1999, pp. 110–111).



One might think of the current discussion about teacher as researcher as a more granular manifestation of lesson study as discussed in the later 1990s and early 2000s. PLCs that have engaged in lesson study would likely have easy transition to conducting teacher-designed experiments focused on specific instructional strategies.

Productive collaboration and spread of effective practice at the individual building level is still insufficient for achieving truly sustainable system improvement. For real system improvement, we look to between-school networks—schools helping schools to amplify the benefits and results of effective collaboration. The between-school collaboration approach has been rising in popularity, not as an alternative to school-based PLCs, but as a complement.

### Professional Learning Networks

Leveraging the evidence that teacher collaboration leads to improved student and school outcomes (Borko, 2004; Darling-Hammond, 2010; Vescio, Ross, & Adams, 2008), more schools, system leaders, and policy makers around the globe are investing in forms of between-school collaboration to further promote and enhance these desired outcomes (Briscoe et al., 2015; Harris & Jones, 2017; Poortman & Brown, 2018). Networking with other schools and colleagues (e.g., teachers from nearby schools with similar demographics, researchers with teachers) expands the breadth and circulation of evidence-based and practice-grounded knowledge, strategies, and tools that educators can access and use (Lai & McNaughton, 2018).

**Types of PLNs.** The designs for between-school collaboratives vary in purpose, form, and function. For example, networked improvement communities (NICs) are popular in the U.S and use improvement science methods adapted from healthcare quality improvement to test and scale evidence-based solutions for addressing critical problems of practice (Bryk, Gomez, Grunow, & LeMahieu, 2015). Also gaining popularity in the U.S. are research-practice partnerships, typically collaborations involving one or more schools or districts and research institutions that are intended to stimulate educator learning and school system improvement by increased use of context-specific data and evidence (Barton, Nelsestuen, & Mazzeo, 2014; Coburn, Penuel, & Geil, 2013; Desimone, Wolford, & Hill, 2016; Farrell et al., 2017; Farrell et al., 2018; Henrick, Cobb, Penuel, Jackson, & Clark, 2017; Muñoz & Rodosky, 2015). Other examples of PLNs include knowledge mobilization networks (found in Canada) that are intended to “promote turning evidence-based research into practice” to achieve education improvement (Briscoe et al., 2015; Ng-A-Fook, Kane, Butler, Glithero, & Forte, 2015), and research learning networks (found in England) which are collaborative networks focused on scaling research-informed teaching practice (Brown & Flood, 2020).

These various forms of between-school collaboration are commonly clustered in the school improvement literature under the construct of PLNs. Such networks involve individuals engaged in collaborative learning with others outside of their regular community of practice (e.g., one or more schools, teachers and university researchers) to improve teaching and learning across the school system (Brown & Poortman, 2018).

**Evidence base and conditions for success.** PLNs are not just a popular school improvement approach. An emerging, and growing, evidence base points to their promise for strengthening teacher engagement and confidence (Owen, 2015; Regilman & Ruben, 2012; Stoll, Bolam, McMahon, Wallace, & Thomas, 2006; Vescio, Ross, & Adams, 2008), teacher instructional practice (Borko, 2004; Dogan, Pringle, & Mesa, 2015; Manfra, 2019; Stoll et al., 2006; Vescio et al., 2008), and student learning (Goddard, Goddard, & Tschannen-Moran, 2007; Manfra, 2019; Stoll et al., 2006; Vescio et al., 2008).

Merely forming a PLN does not guarantee intended results. A variety of conditions can hinder PLN success. For example, changes in the external environment (e.g., funding, leadership turnover) can

threaten the long-term viability of PLNs (Hubers & Poortman, 2018), as can too narrow a focus or a direction and activities that do not fully resonate or directly align with members' pressing needs (Sims & Penny, 2015).

Conversely, several factors tend to support PLN success, such as providing “structured, supported, and properly resourced” opportunities for professional collaboration (Harris & Jones, 2017, p. 22). Based on our experience and informed by current PLN literature (Barletta, et al., 2017; Briscoe et al., 2015; Hargreaves, Parsley, & Cox, 2015; Harris & Jones, 2017; Holdsworth & Maynes, 2017; Poortman & Brown, 2018; Schildkamp, Nehez, & Blossing, 2018; Stoll et al., 2006), we have identified the following PLN enabling conditions:

- Shared vision, goals, and focus
- Collective leadership
- Resources and sustainability supports
- Structured collaboration
- Scientific inquiry and improvement
- Learning and knowledge circulation

*Shared vision, goals, and focus.* It is critical that PLNs identify a network purpose and outcomes that members find meaningful and worthy of their effort. This vision helps define what members hope to gain from participation and what they hope to achieve as a result.

*Collective leadership.* PLNs both require and help build shared leadership. They rely on defined structures, processes, and people to set a strategic direction, ensure ongoing management and development of the PLN, and support the day-to-day work of the network. PLN leadership typically involves a combination of formal roles (e.g., steering committee) and informal opportunities (e.g., breakout session presenter). The majority of leaders should come from within (i.e., network members). In addition, for the work to be prioritized and achieve scale, it is important to have leaders contributing from every level—classroom teachers, administrators, other partners.

*Resources and sustainability supports.* A variety of resources are needed to support and sustain a network's infrastructure and member participation, including financial, human, and technological resources. For example, networks need resources to support in-person convenings as well as intermediary support to coordinate and manage PLN activities and to support member engagement and follow through. Schools and districts also need to find ways to provide time and other resources (e.g., release time, substitute teachers, travel) for teachers to carry out their collaborative work.

*Structured collaboration.* It is important for PLNs to establish collaboration structures, processes, activities, and commitments that are aligned to member needs and network goals and that scaffold collaborative learning, problem solving, and co-creation. Structured collaboration promotes *action*. Examples of supportive structures include: communication routines to promote member interaction; defined network roles (e.g., team facilitators); appropriate collaborative activities (in-person and/or online; frequency of touchpoints); tools to support collaboration (e.g., change frameworks, discussion and data use protocols); and expectations and mutual accountability for member participation.

When networks aligned their structural components with action, they appeared to have further geographical reach, more outputs, an increased number of partnerships, and possibly a greater impact in terms of mobilizing research-based evidence into practice. Specifically, alignment is more than just the existence of network structures and their processes; alignment refers to the ways in which network members come together to

create a synergy that moves the network towards achieving its goals (Briscoe et al., 2018, p. 30).

*Scientific inquiry and improvement.* To promote meaningful action that leads to change, many PLNs incorporate a central focus on teacher research, systematic data use, and/or cycles of inquiry and improvement. Network members engage in contextually relevant, data-informed cycles of professional inquiry that are aligned to local continuous improvement goals in order to enhance classroom practice and learning. Members work together to learn and tailor research-based strategies and approaches to their specific contexts; experiment with these teaching practices in the classroom; gather local evidence to evaluate the results or intervention effects; reflect, learn, and make instructional adjustments; and repeat the cycle.

To reap the benefits of authentic, teacher-led collaboration, using systematic inquiry is a key mechanism for teams of teachers working in school-level PLCs and between-school networks. Disciplined, reflective teaching practice anchored in sound professional judgment and supported by locally derived evidence promotes change that is internally directed rather than externally imposed.

Teams of educators can attain these high-quality attributes by engaging in systematic investigations into teaching practices, which can take the form of action research, inquiry, lesson study, and improvement science—all slightly different approaches to continuous improvement where educators identify situations where results depart from expectations and then use various forms of structured investigation to understand why and make refinements (Barletta et al., 2017, p. 4).

*Learning and knowledge circulation.* Effective PLN collaboration enables individual teacher *and* group learning, or what Hargreaves and Fullan (2012) define as professional capital. Networks thrive when new knowledge, tools, ideas, and strategies are circulated throughout the PLN and beyond. Using PLNs to achieve system-wide change requires ongoing, coordinated co-construction and circulation of new knowledge among individuals and collaborative teams that transfer to the school system level. But to achieve scaled effects, network learning and knowledge circulation must extend beyond dissemination. PLN participants need to engage in parallel activities within their own schools, so that colleagues can use and inform new knowledge gained about school and classroom improvement (Brown and Poortman, 2018).

Although each of these conditions is important on its own, PLNs benefit from establishing them as an interdependent, mutually reinforcing set of conditions. Doing so requires PLN leaders to not only focus on the day-to-day activities and desired outcomes of the PLN, but also nurture and strengthen these enabling conditions over time.

#### Teachers as Researchers within PLCs and PLNs

Because instructional improvement happens within the context of classrooms and schools, we argue that the new paradigm of teacher as researcher can directly enhance PLCs and PLNs. Contextual factors that impact instructional improvement include student and community characteristics (e.g., mobility, languages), school characteristics (e.g., size, locale), and improvement initiatives already underway (e.g., personalized competency based approaches, multi-tiered systems of support, response to intervention).

Using a teacher as researcher approach can provide structure to enhance two of the conditions necessary for PLC and PLN success: structured collaboration and scientific inquiry. Teacher-designed experiments can make the work of PLCs more meaningful by providing a mechanism for teachers to take action on immediate and pressing needs in their school. For example, in a school with a recent influx of English learners, teachers as researchers can test and iterate instructional strategies to support these students right away. The short-cycle nature of teacher-led research can reduce potential lags in

adjusting instruction to meet the needs of special populations. Further, using teacher-led research to support ongoing school initiatives, such as response to intervention, can support and enhance implementation without adding activities to teachers' daily work.

Within PLNs, collaboration centered around teacher-led research can make network activities meaningful across different contexts and promote long-term network sustainability. Highly advanced PLNs organize effectively to create common curriculum materials, engage in lesson study, co-teach, connect students in between-school projects, and so on. But it typically takes concerted effort, time, and commitment for teacher teams within PLNs to get to this level of professional collaboration. A focus on teacher-led research can add significant value in mature networks. At the same time, it can fill a need for new or developing networks. By embedding teacher-led research as a common practice within the network, teachers can benefit right away. As they experience success and build relationships with colleagues, teams can take what they are doing and learning from teacher-led research even further.

Incorporating teacher-led research can also help overcome other common network challenges, such as member turnover and follow through. Unlike collaborative endeavors such as cross-school projects or co-teaching, teams using a teacher as researcher approach are not entirely dependent on others in the group to receive benefit. Teacher-led research can increase team efficiency by allowing members to focus their attention on identifying common problems of practice and deciding how they want to organize their joint work. It can also help teams acclimate teachers who join at different points in time, a common reality that between-school networks face. Embedding a rigorous teacher-led research process can help strengthen PLNs by providing common processes, tools, and routines to smooth the way for teams to do deep inquiry work while leveraging, not impeding, educator needs and autonomy. Repeated use of the same teacher-led research process helps teams develop common language and ways of working together, which further supports team development over time and helps new members more easily join the work.

Strengthening the process of collaboration itself is important but insufficient for achieving a PLN's goals. Using a common approach to scientific inquiry helps PLN members translate good collaboration into tangible action to improve teaching and learning. By using a systematic approach to conduct classroom experiments, teachers generate local, real-time data to rigorously study the use of research-based strategies with their specific students.

Incorporating teacher-led research as a central organizing feature provides the added benefit of helping PLNs overcome a common, significant challenge related to PLN sustainability—generating data and evidence directly from collaborative work that can shed light on the value and impact of network activities. The data teachers generate through their research are not only immediately relevant and useful for informing day-to-day practice. These data can also be aggregated to help network leaders and funders evaluate network-wide outcomes and assess the relative value of resources dedicated to PLN collaboration. Studying and circulating results from teachers' short-cycle experiments can also promote collaborative learning around evidence-based instructional strategies and help prevent ineffective practices from spreading within the network. Circulating knowledge within the network about how specific strategies work with which students under what conditions helps scale-up PLN impact.

In summary, the new paradigm for teacher as researcher legitimizes classroom research and re-establishes educators as key agents of change. This approach to research is amplified when embedded in the PLC and PLN collaboration.

## References

- Barletta, B., Comes, D., Perkal, J., Shumaker, R., Wallenstein, J., & Yang, B. (2017). *Networks for school improvement: A review of the literature*. New York, NY: Columbia University Center for Public Research and Leadership.
- Barton, R., Nelstuen, K., & Mazzeo, C. (2014). Addressing the challenges of building and maintaining effective research partnerships. *Lessons Learned*, 4(1), 1–6. Retrieved from <https://educationnorthwest.org/resources/addressing-challenges-building-and-maintaining-effective-research-partnerships%E2%80%933lessons>
- Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher*, 33(8), 3–15.
- Briscoe, P., Pollock, K., Campbell, C., and Carr-Harris, S. (2015). Finding the sweet spot: Network structures and processes for increased knowledge mobilization. *Brock Education Journal*, 25(1), 19–34.
- Brown, C., & Flood, J. (2020). The three roles of school leaders in maximizing the impact of professional learning networks: A case study from England. *International Journal of Educational Research* 99, 1–10. Retrieved from <https://www.sciencedirect.com/science/article/pii/S088303551931691X>
- Brown, C., & Poortman, C.L. (2018). *Networks for learning: Effective collaboration for teacher, school and system improvement*. London: Routledge. doi.org/10.4324/9781315276649
- Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. G. (2015). *Learning to improve: How America's schools can get better at getting better*. Cambridge, MA: Harvard Education Press.
- Cherasaro, T. L., Reale, M. L., Haystead, M., & Marzano, R. J. (2015). *Instructional improvement cycle: A teacher's toolkit for collecting and analyzing data on instructional strategies* (REL 2015–080). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Central. Retrieved from <http://ies.ed.gov/ncee/edlabs>.
- Coburn, C. E., Penuel, W. R., & Geil, K. E. (2013). *Research-practice partnerships: A strategy for leveraging research for educational improvement in school districts*. New York, NY: William T. Grant Foundation. Retrieved from <https://eric.ed.gov/?id=ED568396>
- Cuban, L. (1992). Managing dilemmas while building professional communities. *Educational Researcher*, 21(1), 4–11.
- Danielson, C. (2011). *The framework for teaching evaluation instrument*. Princeton, NJ: Danielson Group.
- Danielson, C. (2013). *The framework for teaching evaluation instrument*. Princeton, NJ: Danielson Group.
- Darling-Hammond, L. (2010). *The flat world and education: How America's commitment to equity will determine our future*. New York: Teachers College Press.
- Desimone, L. M., Wolford, T., & Hill, K. L. (2016). Research-practice: A practical conceptual framework. *AERA Open*, 2(4), 1–14. doi:10.1177/2332858416679599
- Dogan, S., Pringle, R., & Mesa, J. (2016). The impacts of professional learning communities on science teachers' knowledge, practice and student learning: A review. *Professional Development in*

- Education*, 42(4), 569–588. Retrieved from [https://ir.uwf.edu/islandora/object/uwf%3A22771/datastream/PDF/download/uwf\\_22771.pdf](https://ir.uwf.edu/islandora/object/uwf%3A22771/datastream/PDF/download/uwf_22771.pdf)
- DuFour, R., & Eaker, R. (1998). *Professional learning communities at work*. Bloomington, IN: Solution Tree Press.
- DuFour, R., DuFour, R., & Eaker, R. (2008). *Revisiting professional learning communities at work: New insights for improving schools*. Bloomington, IN: Solution Tree Press.
- DuFour, R., DuFour, R., Eaker, R., & Many, T. (2010). *Learning by doing: A handbook for professional learning communities that work* (2nd ed.). Bloomington, IN: Solution Tree Press.
- Eppley, K., Azano, A. P., Brenner, D. G., & Shannon, P. (2018). What counts as evidence in rural schools? Evidence-based practice and practice-based evidence for diverse settings. *The Rural Educator*, 39(2).
- Farrell, C. C., Davidson, K. L., Repko-Erwin, M., Penuel, W. R., Herlihy, C., Potvin, A. S., & Hill, H. C. (2017). *A descriptive study of IES researcher–practitioner partnerships in education research program: Interim report* (Technical Report No. 2). Boulder, CO: National Center for Research in Policy and Practice. Retrieved from [http://ncrpp.org/assets/documents/RPP-Technical-Report\\_Feb-2017.pdf](http://ncrpp.org/assets/documents/RPP-Technical-Report_Feb-2017.pdf)
- Farrell, C. C., Davidson, K. L., Repko-Erwin, M., Penuel, W. R., Quantz, M., Wong, H., . . . Brink, Z. (2018). *A descriptive study of IES researcher–practitioner partnerships in education research program: Final report* (Technical Report No. 3). Boulder, CO: National Center for Research in Policy and Practice. Retrieved from [http://ncrpp.org/assets/documents/NCRPP-Technical-Report-No-3\\_Full-Report.pdf](http://ncrpp.org/assets/documents/NCRPP-Technical-Report-No-3_Full-Report.pdf)
- Goddard, Y. L., Goddard, R. D., & Tschannen-Moran, M. (2007). A theoretical and empirical investigation of teacher collaboration for school improvement and student achievement in public elementary schools. *Teachers College Record* 109(4), 877–896
- Hargreaves, A., & Fullan, M. (2012). *Professional capital: Transforming teaching in every school*. New York, NY: Teachers College Press.
- Hargreaves, A., Parsley, D., & Cox, L. (2015). Designing rural school improvement networks: Aspirations and actualities. *Peabody Journal of Education: Issues of Leadership, Policy, and Organizations*, 90(2), 306-21.
- Hargreaves, A. (2018). Forward. In C. Brown & C.L. Poortmant (Eds.), *Networks for learning: Effective collaboration for teacher, school and system improvement* (pp. 10-19). New York, NY: Routledge.
- Harris, A. & Jones, M.S. (2017). Professional learning communities: A strategy for school and system improvement? *Wales Journal of Education* 19(1), 16-38.
- Hattie, John. (2008). *Visible learning*. Abingdon, Oxon: Routledge.
- Hedges, L. V., Hedberg, E. C. (2007). Intraclass correlations for planning group randomized trials in education. Northwestern University. Retrieved from <https://www.ipr.northwestern.edu/documents/working-papers/2006/IPR-WP-06-12.pdf>
- Henrick, E. C., Cobb, P., Penuel, W. R., Jackson, K., & Clark, T. (2017). *Assessing research-practice partnerships: Five dimensions of effectiveness*. New York, NY: William T. Grant Foundation. Retrieved from <http://wtgrantfoundation.org/new-report-assessing-research-practice-partnerships-five-dimensions-effectiveness>

- Holdsworth, S. & Maynes, N. (2017). "But what if I fail?" A meta-synthetic study of the conditions supporting teacher innovation. *Canadian Journal of Education*, 40(4), 666–703.
- Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81, 945-970.
- Hord, S. M. (1997). *Professional learning communities: Communities of continuous inquiry and improvement*. Austin: Southwest Educational Development Laboratory.
- Hubers, M.D. & Poortman, C.L. (2018). Establishing sustainable school improvement through professional learning networks. In C. Brown & C. L. Poortmant (Eds.), *Networks for learning: Effective collaboration for teacher, school and system improvement* (pp. 194-204). New York, NY: Routledge.
- Lai, M. K., & McNaughton, S. (2018). Learning networks for sustainable literacy achievement. In C. Brown & C. L. Poortmant (Eds.), *Networks for learning: Effective collaboration for teacher, school and system improvement* (pp. 10-19). New York, NY: Routledge.
- Lewis, C. (2002). Does lesson study have a future in the United States? *Nagoya Journal of Education and Human Development*, 1, 1–23. Retrieved from <https://eric.ed.gov/?id=ED472163>
- Locke, J. (1975). *An essay concerning human understanding*. Oxford, UK: Clarendon Press. (Original work published in 1690).
- Louis, K. S., Marks, H. M., & Kruse, S. D. (1996). Teachers' professional community in restructuring schools. *American Educational Research Journal*, 33(4), 757–798.
- Manfra, M. M. (2019). Action research and systematic, intentional change in teaching practice. In T. D. Pigott, A. M. Ryan, & C. Tocci, (Eds.), *Review of Research in Education*, Vol. 43, 163-196.
- Marzano, R. J. (2017). *The new art and science of teaching*. Bloomington, IN: Solution Tree Press.
- McLaughlin, M. W. (1993). What matters most in teachers' workplace context? In J. W. Little & M. W. McLaughlin (Eds.), *Teachers' work: Individuals, colleagues, and contexts* (pp. 79–103). New York: Teachers College Press.
- Meta-Analysis Database of Instructional Strategies. Retrieved from <https://www.marzanoresources.com/research/database>
- Muñoz, M. A., & Rodosky, R. J. (2015). School districts as partners in research efforts. *Phi Delta Kappan*, 96(5), 42–46. doi:10.1177/0031721715569469
- Ng-A-Fook, N., Kane, R. G., Butler, J. K., Glithero, L., & Forte, R. (2015). Brokering knowledge mobilization networks: Policy reforms, partnerships, and teacher education. *Education Policy Analysis Archives*, 23(122). <http://dx.doi.org/10.14507/epaa.v23.2090>
- Owen, S. M. (2015). Teacher professional learning communities in innovative contexts: "Ah hah moments," "passion" and "making a difference" for student learning. *Professional Development in Education*, 41(1), 57–74. DOI: 10.1080/19415257.2013.869504
- Pine, G. J. (2009). *Teacher action research: Building knowledge democracies*. Thousand Oaks, CA: Sage Publications.
- Poortman, C. L. & Brown, C. (2018). The importance of professional learning networks. In C. Brown & C. L. Poortmant (Eds.), *Networks for learning: Effective collaboration for teacher, school and system improvement* (pp. 10-19). New York, NY: Routledge.



- Raudenbush, S. W. (2005). *Learning from attempts to improve schooling: The contribution of methodological diversity*. Ann Arbor, MI: Applying Multiple Social Science Research Methods to Educational Problems.
- Rosenholtz, S. J. (1991). *Teachers' workplace: The social organization of schools*. New York: Teachers College Press.
- Rubin, D. B. (1974). Estimating causal effects of treatment in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66, 688-701.
- Rubin, D. B. (1977). Assignment of treatment group on the basis of covariance. *Journal of Educational Statistics*, 2, 1-26.
- Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *Annals of Statistics*, 6, 34-58.
- Schildkamp, K., Nehez, J. & Blossing, U. (2018). From data to learning: A data team professional learning network. In C. Brown & C.L. Poortmant (Eds.), *Networks for learning: Effective collaboration for teacher, school and system improvement* (pp. 10-19). New York, NY: Routledge.
- Schneider, B., Carnoy, M., Kiplattrick, J., Schmidt, W. H., & Shavelson, R. J. (2007). *Estimating causal effects using experimental and observational designs*. Washington D. C.: American Educational Research Association.
- Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. Surrey, England: Ashgate.
- Sims, R. L., & Penny, G. R. (2015). Examination of a failed professional learning community. *Journal of Education and Training Studies*, 3(1), 39–45. Retrieved from <https://eric.ed.gov/?id=EJ1054892>
- Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis* (2nd Ed). Los Angeles, CA: SAGE.
- Stigler, J. W., & Hiebert, J. (1999). *The teaching gap: Best ideas from the world's teachers for improving education in the classroom*. New York: Free Press.
- Stenhouse, L. (1975). *Introduction to curriculum research and development*. London: Heinemann.
- Stoll, L., Bolam, R., McMahon, S., Wallace, M., & Thomas, S. (2006). Professional learning communities: A review of the Literature. *Journal of Educational Change*, 7, 221–258. DOI: DOI 10.1007/s10833-006-0001-8
- Vescio, V., Ross, D., and Adams, A. (2008). A review of research on the impact of professional learning communities on teaching practice and student learning. *Teaching and Teacher Education*, 24(1), 80–91.