2016-07-01

A Methodological Analysis of Research into the Effect of Professional Learning Community on Student Academic Achievement

Brandon K. Thacker
Brigham Young University

A Methodological Analysis of Research into the Effect of Professional Learning Community on Student Academic Achievement

Brandon K Thacker

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Doctor of Education

Steven J. Hite, Chair
Julie M. Hite
Scott E. Ferrin
E. Vance Randall
A. LeGrand Richards

Department of Educational Leadership & Foundations

Brigham Young University

July 2016

ABSTRACT

A Methodological Analysis of Research into the Effect of Professional Community on Student Academic Achievement

Brandon K Thacker
Department of Educational Leadership & Foundations, BYU
Doctor of Education

This study analyzed all published research articles examining the relationship between professional learning community efforts (*plc*) (used here as a broader category than, but inclusive of, Professional Learning Communities or PLCs) and student academic achievement (SAA) that reported primary research findings published before January 1, 2015. This study specifically identified primary, quantitative studies of SAA that in context are *plc*, but which may or may not be labeled as such, that were published before January 1, 2015. Analyses examined how many studies of *plc* and SAA were of a descriptive, correlational, causal comparative, quasi-experimental, or experimental design type, evaluated the internal validity of their findings, and assessed the generalizability of each study based on normative expectations of implementation and study design type. Each of the 57 studies meeting the inclusion criteria were evaluated using a Design and Quality of Implementation Matrix. Findings indicated that none of 57 primary research efforts examining *plc* and SAA exhibited acceptable levels of generalizability. For articles demonstrating high design and implementation scores, threats to external validity are presented and discussed. Recommendations are provided for improving the generalizability of research in *plc*.

Keywords: student academic achievement, professional learning community, lesson study, collaboration, methodology, and validity

ACKNOWLEDGMENTS

I am grateful to A. LeGrand Richards, E. Vance Randall, and Scott E. Ferrin for their help in supplying background in educational policy and theory as well as for the time committed to reviewing and improving this dissertation. I am appreciative of the expertise in professional community supplied by Pam Hallam and for the organizational suggestions of Julie Hite which are reflected in the content and flow of this dissertation. To my chair, Steve Hite, who provided insight, critical educational research theory background, and kept the dissertation moving according to schedule, I express deep and heartfelt thanks. I wish to acknowledge the patience of my wife, Shauna Thacker, without whom this project would not have been possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

DESCRIPTION OF STRUCTURE AND CONTENT

This manuscript is presented in the hybrid dissertation format. The hybrid format focuses on producing a journal-ready manuscript which is considered by the dissertation committee to be ready for submission for publication. Therefore, this dissertation has fewer chapters than the traditional dissertation format. The manuscript focuses on the presentation of the scholarly article. This hybrid dissertation includes appended materials such as Appendix A, which contains an extended review of literature, and Appendix B, which is a methods section with elaborated detail on the research approach used in this dissertation project.

The format of the article section follows the *American Journal of Education*'s (AJE) author guidelines (see Appendix C). Citations and references in the article sections follow AJE style requirements rather than APA 6th format. The extended review of literature and detailed method sections follow APA 6th as required by the McKay School of Education at Brigham Young University. AJE is sponsored by Pennsylvania State University, is published by the University of Chicago Press, has a 2014 impact rating of 1.6 and has resided in the top quartile of educational journals since 2006. AJE aims to bridge the methodological diversity of educational scholarship while encouraging vigorous dialog between educational researchers. Articles submitted to AJE are double-blind reviewed by external reviewers. The manuscript length must be under 10,000 words including tables, figures and references. The target audience for the AJE consists of both academics and practitioners in educational leadership.

Background

Overview

The building administrator is the instructional leader for a school, and when a school is perceived to be lacking in student academic achievement (SAA) the building administrator is held accountable and expected to increase SAA. Professional learning community efforts (*plc*), trumpeted as an avenue to raise SAA (Fullan 2002; Lomos et al. 2012; McLaughlin and Talbert 2006; Roberts 2010; Schmoker 2004), have limited generalizability (Jones et al. 2013; Moller et al. 2013). The mixed research on the link between professional learning communities and SAA provides few clues to the instructional leaders as to which studies are reliable, valid, generalizable and hence most profitable to implement at their schools. This study explores the generalizability of primary (original) quantitative *plc* research in print before January 1, 2015.

This study is critical as misapplying the research linking *plc* and SAA risks loss of human and financial capital, institutional credibility, community support, and the educational future of students. A judicious building-level leader will not arbitrarily implement *plc* if there is no strong research base indicating that the risks are justified by the SAA. A school instructional leader finds it difficult to advocate for *plc* on the basis of research claiming increased SAA due to the technical difficulty in determining the generalizability of published research to a particular school context. To empower the instructional leader in better understanding the generalizability of *plc* research, this study poses the following research questions:

1. What primary quantitative research studies of SAA that in context are *plc*, but which may or may not be labeled as such, are available for this study?
2. How many primary studies of *plc* and SAA are of a descriptive, correlational, causal comparative, quasi-experimental, or experimental design type?

3. What is the internal validity of the findings? Does the design of each study and the quality of implementation support the conclusions drawn?
4. What is the external validity, and therefore the generalizability, of each study based on normative expectations of quality of implementation and type of study design?

Need for the Study

An instructional leader, justifying allocation of resources to implement *plc*, must determine the external validity of research linking *plc* and SAA; that is, are the findings of a specific study generalizable to their school? The research on the link between *plc* and SAA varies widely in design type and quality of implementation and, consequently, varies in the internal and external validity of the findings. Many building-level leaders lack the time to determine whether any of the research provides defensible evidence for the role of *plc* in increasing SAA at their site.

The superintendent of schools holds the building leader accountable for the use of scarce resources and requires defensible justification for the leader's use of resources to implement *plc*. Where the application of resources can be justified by recourse to externally valid, generalizable research, the building-leader can, if necessary, defend the decision successfully.

Meta-analyses of *plc* research exist (e.g., Lomos et al. 2011a; Vescio et al. 2008), and building-leaders may initially accept the results as evidence of *plc* generalizability. However, due to the nature of meta-analyses, little information is provided clarifying the generalizability of the individual studies examined. The focus on effect-size in meta-analyses provides little insight into the quality of implementation of the study or the degree to which *plc* were effectively implemented at the site(s) studied. Building-leaders may not be able to discern the potential

limitations of the generalizability of the studies included in meta-analysis research on *plc*. Meta-analysis, in general, assumes that the large sample size will effectively and acceptably balance the limitations of one study with the strengths of another. However, in the Lomos et al. and Vescio et al. meta-analyses, the samples sizes were quite small (n = 7; n = 11). Lacking the covering logic of large sample size, the effect sizes reported could have been produced by studies with poor design or low quality implementation. These two confounding factors matter greatly to the utility of the results in a particular school setting. In contrast to meta-analyses, this study focuses specifically on the type and quality of design implementation in *plc* research, with the goal of evaluating the resulting application to specific school contexts.

Review of Literature

Groundwork for the development of *plc* can be traced to the 1970s and 1980s. When the open classroom teams failed to show expected results (Gamsky 1970), researchers turned to increasing educational effectiveness through professional development (McLaughlin 1979; Runkel et al. 1975). In the 1980s, the concepts of highly effective teams and learning organizations were introduced (e.g., Goodman et al. 1987; Hackman 1980; Kulik and Oldham 1987; Senge 1990). Researchers from various fields identified high work-team differentiation, high integration, performance orientation, and self-government as characteristics of high performing teams (e.g., London and London 1996; Sundstrom et al. 1990). Educational reform efforts would attempt to emulate many of these group characteristics in the decade that followed.

Educational Reform and Professional Development

In 1872, the Japanese government decided to modernize their educational system. Professional developers introduced the *criticism lesson* to expand the educators' instructional repertoire (Mutch-Jones et al. 2012). An aspiring teacher presented an object lesson to fellow

teachers who evaluated the lesson in terms of content, method, and student response. By 1960, the *criticism lesson* had evolved into a format called *lesson study* (Fernandez et al. 2003). In many important ways, Japanese *lesson study* could be considered as *plc*. The term *lesson study* shows up first in American research studies in 1997 (Lewis and Tsucihida 1997; Robinson and Leikin 2012).

Long before 1997, the challenge of Japanese industrial and educational competition had reached America's shores. In response to foreign competition, the Secretary of Education, T.H. Bell published a report entitled, "*A Nation at Risk.*" Recommendation #7 in this report was that "master teachers should be involved in designing teacher preparation programs and in supervising teachers during their probationary years" (The National Commission 1983, 9), signaling that job-embedded professional development was an acceptable avenue toward educational reform. Staff interaction had already received attention as researchers discovered that successful schools exhibited "patterned norms of interaction among staff" (Little 1982, 325).

Collaboration and Educational Reform

Early research on groups of individuals in organizations suggested that decentralization and site-based management were important for effective group development (Murphy 1990). Researchers, seizing upon the insight that "task groups form a link between the individual and the organization" (Gladstein 1984, 499), called for professional development to create communities of educators influencing the direction of reform efforts at the school level (Louis et al. 1998). Multiple strands of collaborative teacher groups developed. Schools moving beyond traditional lab schools by implementing collaborative groups for *both* professional development and educational reform were called *Professional Development Schools* (Darling-Hammond et al. 1995). In 1996, a year prior to the publication of the lesson study research, the National School

Reform Faculty at Brown University developed *Critical Friends Groups*, along with protocols and associated tools intended to help educators improve their practice (Dunne et al. 2000; Key 2006; Nave 2000).

In 1991, Wenger and Lave studied how a new employee is introduced to a set of best practices in a business context, calling such introductions *situated learning* (1991). By 2002, Wenger placed *situated learning* under the heading of *community of practice* (Wenger et al. 2002). Communities of practice, with a focus on continuous learning beyond the introductory period (Lieberman 2009), can become teacher learning communities (TLCs). In Lieberman's TLC, eight to ten teachers at a site agreed to embed formative assessments in their practice. The teachers met regularly to report on their own progress, to discuss the work of other educators, to consider ways to improve the lesson, and to set goals and objectives for the next round. TLCs needed to be supported by administrators, be teacher driven, discuss concerns that emerged from the classroom, and be motivated by a spirit of inquiry (National Council of Teachers of English 2010). In 1990, one year prior to Wenger's work, Peter Senge coined the term, *learning organization* (Senge 1990) to describe an organization that was continually evolving. Senge, writing for business organizations, described the leader's role in a learning organization as that of a teacher. Senge used the of title *teacher* and thereby primed his concept for application to schools. Senge felt that, when applied to education, a learning organization meant redesigning and restructuring the teacher's role.

> There is a huge difference between individual capability and collective capability, and individual learning and collective learning. But this is rarely reflected in the way schools are organized, because education is so highly individualistic. A second dimension of the problem is that educational institutions are designed and structured in a way that

reinforces the idea that my job as a teacher is as an individual teaching my kids. (Senge 1995, 20)

As the prior examples illustrate, many of the reform efforts in the era of collaboration and educational reform were not labeled as PLCs, yet they exhibited many of the essential characteristics of *plc*. This era served as a bridge to the era of *professional community* that led to the era of Professional Learning Communities.

Professional Community and Educational Reform

According to Kruse, the five characteristics of a professional community are: (a) shared values, (b) collaboration, (c) a focus on student learning, (d) de-privatization of practice, and (e) reflective dialogue (Kruse and Louis 1993; Kruse et al. 1995). The inclusion criteria for this study drew heavily from the first three of Kruse's characteristics and combined the fourth and fifth under a PLC's heading of *a reflective focus on results to determine best practices*. The following short discussion demonstrates that first three characteristics of professional community echo the concepts of Senge's learning organization and are shared with TLCs as well.

Shared vision and values. Senge envisioned an integrated system where leaders developed, among other skills, the new skill of building shared vision throughout the company (Senge 1990). When school staff constructed a shared vision for student learning; developed trusting, interpersonal relationships; and embarked upon a program of continuous learning, the staff established professional community (Hord 1997). Shared values provided a foundation for decision making that was open, ethical, collective, and participative (Kruse et al. 1995; Newmann and Wehlage 1995). The focus of the shared vision was on all students learning at high levels (DuFour 2004; Hord 1998).

Collective responsibility for student learning. Educational researchers claimed that professional community aided schools in the development of collective responsibility for student learning (Louis and Marks 1998). "Underlying the earliest discussions of professional community was the core assumption that the group's objective was not to improve teacher morale or technical skills, but to make a difference for students" (Stoll 2007, 3). In schools where collective responsibility for students was high, the research literature claimed student academic gains were higher than in schools were collective responsibility for students was low (Lee and Smith 1996; Lomos et al. 2011b; Moolenaar et al. 2012).

Collaborative efforts focused on student learning. Senge noted that the complex challenges faced by business required collaborative learning among different but equally qualified individuals (1990). Collaboration within a school, Newmann et al. noted, occurred as teachers shared expertise with each other on how to best remediate or reteach concepts not yet mastered by their students (2000). Researchers claimed such sharing influenced students indirectly, with small but positive effects on SAA (Goddard et al. 2007; Newmann et al. 2000). One way teachers collaborated was through de-privatization of practice.

De-privatization of practice. De-privatization occurred when educators viewed each other's teaching and then debriefed their colleagues based on their observations. Team teaching, classroom observations, and peer coaching are examples of this characteristic. De-privatization of practice was the characteristic least likely to be observed in studies of *plc* in schools (Bolam et al. 2005; Lomos et al. 2011b). More frequently teachers de-privatized their student formative data in a team or grade-level effort to increase student achievement in a particular curricular area. The motivation for classroom observation was to allow teachers to trade off the roles of mentor, advisor, and specialist all with an aim of providing aid and assistance to one another (Kruse et.

al. 1993). In view of recent trends towards de-privatized formative class data and the growing rarity of classroom observations, this characteristic was not among the inclusion criteria for this study, being, in a sense, subsumed by the criteria of *collaborative efforts focused on student learning*. If a study noted classroom observations, this fact was noted as a *bonus* characteristic for the research document.

Reflective dialog. Reflective dialog, a characteristic of *plc*, required educators to view their instruction through both the lens of a teacher and the lens of a researcher. As an example, while working with Japanese educators on implementing lesson study into the US curriculum, Fernandez et al. (2003, 173) noted:

> We observed the Japanese teachers continually encouraging the American teachers to see themselves as researchers conducting an empirical examination, organized around asking questions about practice and designing classroom experiments to explore these questions. In particular, the Japanese teachers emphasized four critical aspects of good research: the development of meaningful and testable hypotheses, the use of appropriate means for exploring these hypotheses, the reliance on evidence to judge the success of research endeavors, and the interest in generalizing research findings to other applicable contexts.

The concept of the teacher using a researcher lens is embodied in the fourth criteria for a documents inclusion in this study: a reflective focus on results to determine best practices. This inclusion criterion, with roots going back to the lab school efforts at the University of Chicago, is drawn from the PLCs era.

Professional Learning Communities (PLCs)

Later researchers studied other potential characteristics of *plc* beyond the five originally proposed by Kruse in 1993. One of these potential characteristics constitutes the fourth inclusion

criteria for this study. Others were noted by this study as bonus characteristics of the *plc* research documents.

Continuous learning/formative data. When collaboration moved beyond professional development focused on a set of teacher skills, was continuous, and changed the culture of the school, that professional development was called a Professional Learning Community (PLC) (Louis 2006; Stollar 2014). Changing a school culture in this way required time and multiple iterations of practice (Louis 2006). In a PLC, teachers would strive to develop not only instructional strategies but also the inclination to continually improve their instructional prowess (DuFour et al. 2006). Feedback between the students and teacher nurtured the continuous learning process. Using data from common formative assessments was proposed as a critical PLC component in 2004 (Dufour 2004). *A reflective focus on results to determine best practices* was used as the fourth inclusion criteria for this study. Beyond these critical four criteria, two additional *plc* characteristics were noted in the literature, although they were not required for inclusion in this study.

Shared leadership. Hord contended that adding shared leadership to Kruze's characteristics of a professional community created a PLC (Hord 1997; Hord 1998). PLCs expected those closest to the instruction to control the allocation of resources directly affecting instruction (Sackney and Walker 2005). Mitchell argued "it means that leadership is enacted throughout the school by a variety of individuals in a variety of ways" (Mitchell and Sackney 2001, 2).

High trust/supportive structures. As teachers honed their skills collaboratively to improve student instruction, they developed a greater trust in one another and in the students (Louis and Marks 1998). Trust must also exist between teachers and administrators sufficient for

teachers to feel comfortable asking for help and exposing weaknesses (Byrk and Schnieder 2002). Trust and PLCs were seen as self-reinforcing, creating additional levels of trust to allow for advanced levels of community (Huffman and Kalnin 2002). In order to foster high levels of trust, appropriate temporal and social structures were needed to encourage PLC success (Stoll and Louis 2007). Adequate time and proximity were listed as critical components for effective PLCs (Bolam et al. 2005). For example, one study on social capital and professional community networks found the greatest variance in student achievement hinged on the proximity of veteran and novice teachers in learning teams (Penuel et al. 2009).

Essential Characteristics of *plc*

A *plc* is a complex phenomenon (Wilson 2014), and debate rightly continues over what constitutes its critical components thereby avoiding the "cold comfort of final definition" (Clegg et al. 2005, 149) and allowing continued insights into the restructuring of teacher roles. The following eight characteristics of an effective PLC were presented to the United Kingdom's Department of Education and Skills in 2005, following a nationwide survey of PLC efforts (Bolam et al. 2005):

1. A shared value and vision
2. Collective responsibility for pupil learning
3. Learning focused collaboration
4. Individual and collective professional learning
5. Reflective teaching
6. Partnerships and shared authority
7. Inclusive membership, high trust levels
8. Supportive structural adaptations

Wilson notes Bolam's claim that the most common characteristic of a PLC was shared mission and values. The next three most frequently noted characteristics were collective responsibility for pupils' learning, collaboration focused on student learning, and reflective professional inquiry (Bolam 2005; Wilson 2014).

In the *Method* section I outline the inclusion criteria for a document to be included in the census of articles on *plc* and SAA. The first three criteria are essentially the same ones listed by Wilson with the fourth being modified to reflect DuFour's contention that PLCs must be data driven, with formative data driving intervention and remediation strategies (Dufour 2004; Wilson 2014).

Internal and External Validity

A building-leader viewing data from a *plc* study can make inferences from the data. Validity measures the quality of an inference. Building leaders examining data from *plc* research frequently make two types of inferences. One inference might be that the *plc* are responsible for the gains/lack of gains in the research. Such an inference is made about the results internal to the research context and the quality of this inference is called *internal validity*. A second inference might be that the findings of gains/lack of gains would occur at the building-leader's site. This inference refers to the generalizability of results outside of the research context, and the quality of this inference is called *external validity*. The ideal *plc* research will be conducted and designed in such a way as to provide high confidence in making both types of inferences from the data. While the main focus of this study is the generalizability of the research results to the building-leader's site (external validity), high confidence should also exist that *plc* caused the reported findings (internal validity) since the building leader will implement *plc* in expectations of student gains.

Factors that reduce internal validity are usually classified under headings such as ambiguous temporal precedence, selection bias, history, maturation, pre-test/post-test, instrumentation change, mortality, regression to the mean, and diffusion effects. Threats to external validity are usually classified under headings such as reactivity and experimental effects (Campbell and Stanley 1963; Campbell and Stanley 1984). For this study, only those research efforts that allowed a high level of generalizability were deemed to have findings with external validity.

Method

Many reviews of *plc* and SAA limited their focus to those efforts explicitly labeled Professional Learning Communities (upper case PLC) and/or Professional Community (e.g., Lomos et al. 2011a; Yoon et al. 2007). A broader set of search parameters allowing *critical friends groups* and *communities of practice* used in another study found only 11 papers qualified for review (Vescio et al. 2008) and eight of the 11 dealt with SAA. While such a limited focus creates a tidy sample of studies to be readily accessed and reviewed, labels can also be overly restrictive and potentially misleading. For example, some reform efforts labeled PLC lack the cooperative characteristics claimed by experts to identify *bona fide* PLCs. Other reform efforts, under different labels, exhibit many of the essential cooperative characteristics and perhaps additional characteristics beyond those in a strictly identified PLC. As with some other instructional practices, "professional learning communities (PLCs) and teacher learning teams (LTs) can be traced to many sources" (Gallimore et al. 2009, 538) and exhibit varying sets of characteristics.

This analysis of *plc* research was based on a census of quantitative, primary research articles in English-language scholarly journals published between January 1, 1980, and January

1, 2015. After making a determination as to the design type of the research described in the article, this study next analyzed the likelihood that the *plc* created the student gains based on typical design expectations and limitations regarding causal claims (Jackson 2015; Mertens 2015; Mills and Gay 2015).

The research articles included in the census for this study had to both contain published quantitative, primary results concerning *plc* and SAA and exhibit, at least, the four following *plc* characteristics found most frequently to be part of *plc* (Bolam et al. 2005; DuFour 2004; Wilson 2014):

1. shared vision and values,
2. collective responsibility for and focus on student learning,
3. collaborative teaming and learning efforts, and
4. a reflective focus on results to determine best practices.

This census included all efforts identified as *plc* based on the above inclusion criteria, even those not identified specifically as professional learning communities. Rather than a handful of studies that qualified (e.g., Yoon et al 2007) this census resulted in 57 research articles. The decision to use a broad definition of *plc* (along with the use of the lower case acronym for such efforts) was purposeful and facilitated the following advantages over previous efforts: (a) no effective criticism can be leveled that the outcome of the analysis was predetermined or manipulated toward any particular objective by excluding favorable or unfavorable studies from the census through restrictive inclusion criteria, (b) no data from the more constrained studies has been lost to this analysis since a study qualifying under a more rigorous definition of *plc* would be included in this analysis, and (c) there is decreased chance of

parochial or cultural misinterpretation of the findings since there was no set of trademarked buzzwords used to determine inclusion.

Over 200 potentially qualifying articles were selected from searches on the Internet, EBSCO database, Web of Science database, or ERIC. The online searches used the following keywords, in various combinations: Achievement, Gains, Professional Community, PLC, Learning Community, Professional Learning, Lesson Study, Multi-Tier Systems of Support, Response to Intervention, Teacher Learning Community, Teacher Networks, Communities of Practice, and/or Critical Friends Group. Studies not looking at quantitative measures of SAA as a dependent variable, or containing no primary student achievement data, were eliminated. The remaining articles were examined to see if the professional community efforts described qualified them for inclusion as *plc* by exhibiting at least the four *plc* characteristics in the inclusion criteria, resulting in 57 articles in the census.

Each of the 57 qualifying articles was evaluated using an adaptation of Reynold's design and implementation evaluation matrix, itself based on Hite's checklist for reviewing research documents (Hite 2001; Reynolds 2005). Study design components were marked as either being present in the article (Yes) or not found (No). The quality of the implementation of the corresponding components of the design were marked with a 1, 2, or 3 (Low, Medium, or High) based on expectations put forth in typical educational research design texts (e.g., Creswell 2014; Gay et al. 2012; Mertens 2015). Note that all scores were based on published details in the collected articles only. See Table 1 for a condensed version of the Design and Implementation Evaluation Matrix. Actual matrices used in the original analysis are available upon request from the corresponding author.

Table 1

Design and Implementation Evaluation Matrix

Design Element	**Yes/No**
Purpose and problem(s)	Purpose found and noted? Problems found and noted? Variables specified? Context of research clarified?
Research method and methodology	Methods presented and discussed? Methodology described (descriptive, correlational, causal-comparative, quasi-experimental, experimental)
Quality of Implementation	**Low/ Medium/ High** 1 / 2 / 3
General Procedures	Procedures sufficiently detailed? Procedures reasonable for study?
Sampling	Population defined? Randomly selection?
Instrumentation	Instrumentation reliable/valid? Instrumentation appropriate for context? Training of researchers adequate?
Data Analysis	Analysis suitable for methodology? Statistical information complete? Data are appropriate to context? Statistical significance considered?
Conclusions	Internal validity? External validity? Conclusions warranted?

As an example of how the matrix was used, one can consider the implementation component *Sample selected randomly*. If the research study employed some type of standard random sampling, such as simple random or stratified proportional random sampling, a mark of High (3 points) was awarded. If the research employed a convenience sample or a purposive

sampling method (e.g. “three schools are great examples of *turn around* schools and these schools are the focus of our study”), a mark of Low (1 point) was awarded. A mark of Low (1 point) was also awarded when a low response rate countered the design value of an attempted random sampling. If the research employed a randomized *cluster* sampling method or some other attempt to apply a random sampling procedure that masked the value of randomly selecting participants, a value of Medium (2 points) was awarded. Each sampling method has benefits and disadvantages, and it is the fit of the sampling method to the context that is most critical in determining validity (Mills and Gay 2015; Hite 2001).

Several examples will help clarify. When the sampling was random, design of research controlled for confounding variables, quality of implementation was high, and baseline data was sufficient to ensure that the *plc* was the cause of the reported student achievement, the article received a 3 for the first (internal) validity question. When the research design was quasi-experimental, quality of implementation was high, and comparative data was used to sufficiently combat all potential threats to validity, the article received a 3 for both (internal and external) validity questions in the Conclusions box. When threats to validity were not sufficiently counteracted when using a causal comparative study design, the research received a 1 or a 2 for both validity questions but might, if conclusions were carefully drawn by the researchers, receive a 3 on Conclusions being warranted.

Findings

The first section provides findings regarding the number of articles found to meet the inclusion criteria and the design type of each (research questions 1 and 2). The second section presents data on whether the quality of implementation and the study design provide confidence

in the article's conclusions (the internal validity of the findings). Lastly, the question of generalizability (the external validity) of the articles' conclusions is examined.

Design Type of Census Articles

Table 2 provides the findings on how many primary studies of *plc* and SAA were of a descriptive, correlational, causal comparative, quasi-experimental, or experimental design type. The table addresses both research questions 1 and 2 regarding the published articles contained in the census of studies.

Table 2

Census Articles by Study Design Type

Design Type	Number Found	Percent of Census
Descriptive	17	30
Correlational	19	33
Causal Comparative	18	32
Quasi-experimental	3	5
Experimental	0	0
Total	57	100

There are a comparable number of studies which are descriptive, correlational or causal comparative in nature. Such uniform distribution does not extend to the quasi-experimental or experimental designs, which will be discussed later in the findings section.

Internal Validity: Design Type and Quality of Implementation

This section describes, among other things, how well the implementation fit the design type of the study. While it is impossible to know the author's intended design type, the steps taken to implement the study and presented in the article are taken as evidence of the intent of

the author. The numeric rating of the implementation of the study should yield high scores for those studies with internal validity. It is reasonable to seek for internal validity since proponents of *plc* claim an effect on SAA and therefore must design and implement their study in such a way as to convince the reader that the change in SAA is due to *plc* and not due to other factors.

None of the studies in the census demonstrated high internal validity. This finding does not mean that *plc* never affect SAA, but based on the research one must remain agnostic about the impact. This paper adopted a broad definition of *plc* and included five to seven times the number of articles in the census as past researchers, so the finding came as a surprise to this author. Hereafter is a discussion of the internal validity of the *top ten* scoring studies to illuminate why no article in the census indicated a high confidence that *plc* created the student gains noted in the specific contexts of the research sites.

Five (50%) of the *top ten* studies (Bolam et al. 2005; Daly et al. 2011; Lomos et al. 2012; Moolenaar et al. 2012; Stahl et al. 2013) were correlational in design, and their findings dealt with association and not causation (see Table 3). Studies that are correlational in design do not attempt to determine causation. A correlation study might provide some insight into how *plc* might interact with SAA but provides only low to medium confidence that *plc* caused the changes in SAA.

The 16 case study schools (volunteers) in Bolam's (2005) report to the British educational department on PLCs were presented as correlations and made no claims beyond association. Lomos et al. (2011b) found small effect sizes for schools with higher levels of *plc*, and included these warnings, "After controlling for important variables at student and teacher/school levels, an additional 7% of the variance among schools was explained by the presence of the five characteristics of professional community, with an effect

Table 3

Implementation Scores for Correlational Study Designs

Article	Bolam 2005	Daly 2011	Lomos 2012	Moolenar 2012	Stahl 2013
Quality of Implementation	1/2/3	1/2/3	1/2/3	1/2/3	1/2/3
General Procedures					
Procedures detailed	3	3	3	3	3
Procedures reasonable	3	2	3	3	3
Sampling					
Population defined	3	3	3	3	3
Sample selected randomly	1	2	1	1	1
Measurement					
Instrumentation reliable/valid	3	2	3	2	2
Reasonable for context	3	3	3	3	3
Training adequate	2	2	3	3	3
Data analysis					
Analysis suitable	3	2	3	3	3
Statistical information	3	3	2	3	3
Data use appropriate	3	3	3	3	2
Stat. significance considered	3	3	3	3	3
Results					
Purpose of each problem addressed	3	3	3	3	3
All information in original	3	3	3	3	3
Conclusions/generalizability					
Internal validity high	2	2	2	2	2
External validity high	1	1	1	1	1
Conclusions warranted	3	2	3	3	2
Total Implementation Points	43	39	44	42	40

size of .25, which could be considered small" (729) and "The second limitation of this study lays in measuring student achievement at one point in time, which did not allow [us] to establish whether strong professional communities lead to higher levels of achievement and vice-versa" (729). Daly's examination of human capital and student achievement likewise found no significant relationship between *plc* and SAA (2011). Daly ends with this disclaimer, "our analysis strategy was not intended to determine causality" (2011, 27). Lomos et al worked with Dutch secondary students' exit scores and four categories of PLCs but found no significant relationship between PLCs and SAA, noting only one dimension had a positive effect: "The *reflective dialogue* sub-dimension was the only one significantly and positively associated with student achievement, with an effect size of .24" (2012, 123), an effect size that a year earlier Lomos had considered small.

Moolenaar et al.'s research (2012) into networks and teacher collective efficacy found that "a direct effect between advice network characteristics and student achievement could not be evidenced" and is similarly correlational in design. Stahl's (2013) mixed methods Response to Intervention (RTI) research employed purposive sampling, was correlational in nature, and found comparable growth rates between the experimental and the comparison groups. Each of the high scoring correlational studies provided medium confidence that the *plc* was responsible for the student gains noted but did not eliminate other potential factors.

Four of the *top ten* studies (40%) were causal-comparative in design (see Table 4). The four studies (Boaler and Staples 2008; Goddard et al. 2007; Moller et al. 2013; Smylie et al. 2003) attempted to isolate the influence of *plc* on student achievement after the data was collected. Smylie et al.'s report to the Annenberg Foundation on SAA was causal-comparative in design in the beginning but over time failed to disentangle the focus schools from other district

Table 4

Implementation Scores for Causal-Comparative Study Designs

Causal-Comparative	**Boaler**	**Smylie**	**Goddard**	**Moller**
Quality of Implementation	1/2/3	1/2/3	1/2/3	1/2/3
General procedures				
Procedures sufficiently detailed	3	3	3	3
Procedures reasonable for study	3	3	3	3
Sampling				
Population defined	3	3	3	3
Sample selected randomly	1	1	3	1
Measurement				
Instrumentation reliable/valid	3	3	2	2
Instruments reasonable for context	2	3	3	2
Training of researchers adequate	3	3	2	3
Data analysis				
Analysis suitable for methodology	3	3	3	2
Statistical information complete	3	3	3	3
Data used are appropriate	2	2	2	3
Statistical significance considered	3	3	3	3
Results				
Purpose of each problem addressed	3	3	3	3
All information in original	3	1	3	3
Conclusions and generalizability				
Internal validity high	2	2	2	2
External validity high	1	1	1	2
Conclusions adequately warranted	2	3	2	2
Total Implementation Points	40	40	41	40

reform events occurring in the same time period (2003, 48), leaving Smylie's findings of no significant difference in student achievement to be based only on patterns or association.

Boaler (2008) executed casual-comparative research of three schools, purposively selected, noting that the demographics of the schools differed at the beginning of the research study. Two of the three schools served as the comparison group and the third school, Railside, was used as the experimental school. Railside students showed impressive gains over the other two schools. In this five-year study, Railside students spent twice the time on subject as at the other two schools, a fact that likely accounted for some of the observed difference in SAA. The time on subject and the difference in initial demographics were never controlled for in the research, limiting its internal validity.

The Goddard causal-comparative research (2007) selected elementary schools randomly from a purposively selected school district. The Midwestern school district became the focus of their research. The authors themselves noted that a more randomly selected population is required before generalization of results can be meaningful. Within the Midwestern school district, self-reported survey items relating to *plc* were associated with increased SAA. Considering the concerns of using self-reported survey data and the correlational nature of the analysis, a mark of 2 for internal validity was awarded.

Moller's data analysis (2013) drew on *cross-classified growth* modeling theory. Cross-classified growth modeling grows out of hierarchical modeling theory (Raudenbush and Byrk 2002) and shares many of the assumptions of regression analysis. Discussion about the ability of cross-classified growth models to support causality is ongoing (Desimone et al. 2013; Grady and Beretvas 2010). Fielding and Goldstein (2006, 23) hint at the inadequacy of using cross-classified models to indicate causality:

> Such model results on observational data should not be over interpreted to yield firm causal explanations. However, they might suggest ways in which interventions might be designed which can then be trialed in a designed framework to yield more firmly based interpretations.

Moller's database (2013) contained four data points from Kindergarten to eighth grade. Over the nine-year period, from 1998 to 2006, numerous opportunities existed for the educational environment to change (such as the passage of No Child Left Behind in 2001) and thereby threaten the internal validity, since the projections were made in 1998. Moller is to be applauded for applying this technique to a relatively large sample (4,000+ students) and for the careful statistical checks carried out at various points in the data analysis. The correlational nature of Moller's assumptions limits the internal validity of the findings.

Only one (10%) of the *top ten* studies was quasi-experimental in design (Saunders et al. 2009), suggesting high potential for internal validity. However, regarding internal validity, the developers intervened robustly in the experimental schools. They concluded that "when developers themselves deliver the intervention, the effect size is more likely to be overstated" (Saunders et al. 2009, 1027). A quasi-experimental study design may claim generalizability of the study results. Saunders et al. is therefore discussed in the next section on external validity as well.

External Validity: Quality of Design and Implementation

Research using an experimental or a quasi-experimental design can most easily claim external validity. Of the 57 studies, none were experimental in design, with only three (5%) being quasi-experimental studies; of those three studies, only Saunders (2009) received high scores in the Quality of Implementation matrix (see Table 5).

Table 5

Implementation Scores for Quasi-Experimental Study Designs

Quasi-Experimental	**Gallimore (2009)**	**Saunders (2009)**	**Pang (2012)**
Quality of Implementation	1 / 2/ 3	1 / 2 /3	1 / 2 / 3
General Procedures			
Procedures sufficiently detailed	2	3	3
Procedures reasonable for study	2	3	2
Sampling			
Population defined	3	3	2
Sample selected randomly	1	1	1
Measurement			
Instrumentation reliable/valid	2	2	3
Instruments reasonable for context	2	3	3
Training of researchers adequate	2	2	2
Data analysis			
Analysis suitable for methodology	2	3	1
Statistical information complete	1	3	1
Data used are appropriate	3	3	2
Statistical significance considered	1	3	1
Results			
Purpose of each problem addressed	3	3	2
All information in original	2	2	2
Conclusions and generalizability			
Internal validity high	2	2	2
External validity high	2	2	1
Conclusions adequately warranted	2	3	1
Total Implementation Points	32	41	29

Therefore, no studies exhibited external validity. Of the three quasi-experimental studies, both Saunders et al. (2009) and Gallimore et al. (2009) examined the same schools. Saunders et al. (2009), attempted to ensure that the experimental and comparison schools were comparatively similar, even compensating for the mid-project changes in design. Still the authors pointed out two factors limiting the external validity of their findings. First, the schools were self-selected with years of prior exposure to similar reform efforts. Second, it involved only urban, elementary schools with high percentages of Hispanic students. In Gallimore et al. (2009), the self-selected participating schools were also urban elementary schools. In addition, Gallimore's comparison measures were only taken at the beginning of the project (which experienced a major change in research design mid-project). Regarding the internal validity of their study the authors remarked that "we cannot satisfactorily specify the contribution of several plausible alternative mediators" (Gallimore et al. 2009, 542).

The third study (Pang and Ling 2012) used a quasi-experimental design involving four teachers (two in a comparison group and two in the learning study group) in a Hong Kong primary school and presented no statistical validation that the control group was initially similar to the experimental group. The lack of random selection also undermined the generalizability of the findings, while the lack of statistical control undermined the internal validity of the research. However, the sampling information may be missing due to publishing guidelines as the authors stated that "due to space limitations, we highlight some of the important features of this learning study that illustrate how learning study, and variation theory in particular, can contribute to professional development among teachers" (Pang and Ling 2012, 600).

In summary, none of the 57 research studies in the census exhibited high levels of external validity, meaning that their findings cannot reasonably be expected to pertain to

locations beyond the context of the study. The three quasi-experimental design studies lacked randomly selected samples necessary for external validity of the findings. Among the *top ten* studies with high Design and Implementation Evaluation Matrix scores, the four with a casual-comparative design failed to compensate for purposive samples or historicity threats to the findings. While some of the research in the census produced findings reflecting a positive association between *plc* and SAA, those findings did not support external validity and should not produce expectations or inferences of similar results elsewhere.

Discussion

Research questions 1 and 2 dealt with the state of *plc* research. The bulk of the research in the census was descriptive, correlational, or causal comparative in design (95%) and, as such, only lays the foundation for causal studies. Two of the three quasi-experimental research studies, both published in 2009, looked at the same urban, title I schools with high levels of Hispanic students (English Language Learners). The third study, published in 2012, was initially submitted for publication in 2010 and likely reported findings from Hong Kong's 2009 school year. In the six years since 2009, no *plc* research using quasi-experimental design was found. No research studies using a quasi-experimental design have been published on school populations that were rural or suburban.

The highest scoring quasi-experimental effort by Saunders (2009) did not use the terms PLC or professional learning community in its title (*Increasing Achievement by Focusing Grade-Level Teams on Improving Classroom Learning: A Prospective, Quasi-Experimental Study of Title I Schools*) or keywords (Keywords: professional development, school/teacher effectiveness, educational reform, longitudinal studies, elementary schools, organization theory/ change). This

article would likely have been over-looked if search terms had been limited to those for traditional PLCs.

Research question #3, "Does the design of each study and the quality of implementation support the conclusions drawn?" dealt with the internal validity of the studies. No studies were found with high internal validity. Much of this may be due to the large number of studies that were correlational or descriptive in design and no claims were made regarding causation. The dearth of *plc* studies with high internal validity in this census indicates that while many correlational studies hint at an association between *plc* and SAA, causation appears to be a difficult dimension to establish. Even though one-third of the articles in the census were causal-comparative in design, the focus on the linkage to SAA needs to be better established.

Varying levels of fidelity were found in the way *plc* were implemented using various subsets of Bolam et al.'s nine characteristics (2005). While beyond the scope of the research questions for this study, it may be that all nine characteristics are required before *plc* can clearly influence SAA. It may be that an as yet undiscovered additional characteristic is needed.

Research question #4 addressing the generalizability of a study's findings dealt with issues of external validity. Since none of the research designs in the census were experimental and only three were quasi-experimental in design, finding research results with external validity to support generalizability was difficult. Of the 57 research studies in this census, not one evidenced high levels of external validity. In fact, few of the studies even claimed a reasonable amount of external validity, while 14 studies actually disclaimed it. To better assist instructional leaders, future research efforts should seek to design studies that provide reasonably high levels of external validity.

Without clear external validity, the instructional leader, who does not enjoy the luxury of time, may take one of three options in deciding how to respond to the findings of *plc* research. First, given that "causal research is often difficult to locate and therefore its use is rare in educational policy endeavors" (Hite 2001, 60), a building-level administrator may choose to pursue professional learning communities without claiming that the research reasonably warrants expectation of increased SAA, stressing instead other potential and inferred benefits of *plc* (Ermeling and Gallimore 2013; Moller et al. 2013; Sleegers et al, 2013; Voelkel 2011). Second, a building-level administrator may rely on proximal similarity (Campbell and Stanley 1984; Trochim 1998). Proximal similarity is the placing of contexts into their relative ranking of similarity. For instance, if a study looked at subjects similar to one's students, in a setting similar to one's school, with backgrounds similar to one's student demographics, then one can decide that the settings are likely proximally similar to one's site and that one may expect similar results. While one never can generalize with certainty using proximal similarity, one's confidence increases the more that the study context approaches that of one's own school. Third, the building level administrator may pursue alternate reform efforts separate from *plc*.

Limitations

While this census sample reviewed the largest number of research studies on *plc* and SAA to date, some limitations exist both to the census and the conclusions drawn. Limiting the census to studies that published quantitative, primary results concerning *plc* and SAA limited our discussion to studies comparing gains in test scores. All of the limitations of each such assessment to capture SAA automatically apply. This census sample includes only research results published in the English language. This fact removes from the census research published in other languages. Second, electronically indexed articles are more likely to be found in highly

developed countries, though Internet searches helped compensate for this regional bias. Third, the fact that the research is published introduces three potential sources of bias: (a) small, single school, independent efforts by principals to implement professional learning community are not intended for publication and so their results are less likely to be in the census, (b) studies published in educational journals tend to mirror the interests of the journal editor and/or journal audience. Statistically non-significant findings are far less likely to be published, which could slant the census towards research demonstrating statistically significant relationships or those deemed worthy of publication, (c) editors' restrictions on the amount of space describing study methodology may require submitters to eliminate descriptive material that would otherwise be included in the scores on the Design and Implementation Evaluation Matrix.

The need for studies with stringent design constraints, producing high external validity of findings, could be somewhat reduced if the contextual factors most critical in determining proximal similarity were included in the published description of a study. This strategy may stretch the normative limitations on describing research design to clarify design, context, and implementation elements. The additional design and context details would assist the building-level administrator in applying proximal similarity to a school's own context. Additional research is needed to estimate the utility of each contextual factor as it facilitates the building leader in the application of proximal similarity.

Suggestions for Further Research

Future research needs to determine the extent to which *plc* have been implemented on site (Sleegers et al. 2013) using metrics and tools already available (Taylor et al. 2014; Whalan 2012; Wiliams et al. 2007) before attempting to analyze if student academic achievement increased at sites labeled as PLCs. Too frequently, little information, other than self-attributed survey results,

is available to determine how fully a PLC was implemented in the *plc*. Evaluating to what extent the school's faculty bought in and practiced the characteristics of a PLC should not rest solely on a faculty self-evaluation. An objective measure of PLC implementation should be provided to help reduce the influence of teachers' *fog of battle* experience during instruction on the self-evaluation survey results.

Since educational researchers are limited in their ability to randomly place students in *plc* environments, care should be taken to randomize the unit experiencing *plc*. Effort should then be expended to ensure that the selected *plc* units (schools or classroom) and corresponding control units match in regard to demographics, learning time, and educational context. Strategically selected control groups within and outside the same district can compensate for the lack of randomly selected students.

The need for studies with stringent design constraints, producing high external validity of findings, could be somewhat reduced if the contextual factors most critical in determining proximal similarity were included in the published description of a study. This strategy may stretch the normative limitations on describing research design to clarify design, context, and implementation elements. The additional design and context details would assist the building-level administrator to apply proximal similarity to a school's own context. Additional research is needed to estimate the utility of each contextual factor as it facilitates the building leader in the application of proximal similarity.

With the goal of generalizable research that connects *plc* and SAA, a serious need exists for high-quality, casual-comparative or quasi-experimental research. Study designs that avoid purposive samples, instead utilizing random selection of participants, should be encouraged and published.

Imagine a world where instead of spending untold hours reading published research, a building-level administrator can review four or five quality studies and immediately pivot to implementing instructional practices that empower teachers to help students achieve at higher levels. Such a world beckons education researchers to implement quality research designed to maximize the generalizability of the findings to contexts beyond those of the specific research studies. PLCs have come a long way in the 20 years since the professional development schools of 1995. However, the current state of PLC development would be anticlimactic if researchers fail now to ply the available research tools to generate findings with internal validity that support the link between *plc* and student achievement and external validity to support the generalizability of their findings to other educational contexts striving to improve student learning.

Do not imagine that these findings assert that *plc* do not improve SAA. There are simply no primary, quantitative studies that prove that *plc* do so. The existing evidence fails in ways that may not be noted by the voices trumpeting the academic gains of *plc*. This need not stop the implementation of *plc* but let us not assume that research has as yet established a positive, strong link between *plc* and SAA.

References

Boaler, Jo, and Staples, Megan. 2008. ”Creating Mathematical Future through an Equitable Teaching Approach: The Case of Railside School.” *Teachers College Record,* 110 (3): 608-645.

Bolam, Ray, McMahon, Agnes, Stoll, Louise, Thomas, Sally, and Wallace, Mike. 2005. *Creating and Sustaining Effective Professional Learning Communities*. Bristol, UK: University of Bristol.

Byrk, Anthony, and Schnieder, Barbara. 2002. *Trust in Schools*. New York, NY: Russell Sage.

Campbell, Donald, and Stanley, Julian. 1963. *Experimental and Quasi-Experimental Designs for Research*. Chicago, IL: Rand McNally.

Campbell, Donald, and Stanley, Julian. 1984. *Experimental and Quasi-Experimental Designs for Research*. Boston, MA: Houghton Mifflin.

Clegg, Stesart, Kornberger, Martin, and Rhodes, Carl. 2005. “Learning, Becoming, Organizing.” *Organization* 12 (2): 147-167.

Creswell, John. 2014. *Educational Research*. Essex, UK: Pearson Education Limted.

Daly, Alan, Moolenaar, Nienke, Der-Martirosian, Claudia, Canrinus, Esther, and Chrispeels, Janet. 2011. “A Capital Investment: The Effects of Teacher Human and Social Capital on Student Achievement in Improving Schools.” In *International Congress on School Effectiveness and Improvement*, manuscript under review. Lymassoll, Cyprus.

Darling-Hammond, Linda, Bullmaster, Marcella, and Cobb, Velma. 1995. “Rethinking Professional Leadership Through Professional Development Schools.” *The Elementary School Journal* 96 (1): 87-106.

Desimone, Laura, Smith, Thomas, and Phillips, Kristie. 2013. “Linking Student Achievement Growth to Professional Development Participation and Changes in Instruction: A Longitudinal Study of Elementary Students and Teachers in Title I Schools.” *Teachers College Record* 115 (5): 1-46.

DuFour, Rebecca, Eaker, Robert, and DuFour, Richard. 2006. *Learning by Doing: A Handbook for Professional Learning Communities at Work*. Bloomington, IN: Solution Tree.

DuFour, Richard. 2004. What is a "Professional Learning Community"? *Educational Leadership* 61 (8): 6-11.

Dunne, Faith, Nave, Bill, and Lewis, Ann. 2000. “Critical Friends Groups: Teachers Helping Teachers Improve Student Learning.” *Phi Delta Kappan,* 28 (4): 31-37.

Ermeling, Bradley, and Gallimore, Ronald. 2013. “Learning to be a Community: Schools Need Adaptable Models to Create Successful Programs.” *JSD: The Learning Forward Journal* 34 (2): 42-45.

Fernandez, Clea, Cannon, Joanna,and Chokshi, Sonal. 2003. “A US-Japan Lesson Study Collaboration Reveals Critical Lenses for Examining Practice.” *Teaching and Teacher Education* 19 (2): 171-185.

Fielding, Antony, and Goldstein, Harvey. 2006. “Cross-classified and Multiple Membership Structures in Multilevel Models: An Introduction and Review.” In *Research*, Department for Education and Skills. Birmingham, UK: University of Birmingham.

Fullan, Michael. 2002. “The Change Leader.” *Educational Leadership* 59 (8): 16-21.

Gallimore, Ronald, Ermeling, Bradley, Saunders, William, and Goldenberg, Claude. 2009. “Moving the Learning of Teaching Closer to Practice: Teacher Education Implications of School-Based Inquiry Teams.” *The Elementary School Journal* 109 (5): 537-553.

Gamsky, Neal. 1970. “Team Teaching, Student Achievement and Attitudes.” *The Journal of Experimental Education,* 39 (1): 32-45.

Gay, Lorrie, Airasian, Peter, and Mills, G. 2012. *Educational Research*. Chicago, IL: Pearson.

Gladstein, Deborah. 1984. “Groups in Context: A Model of Task Group Effectiveness.” *Administrative Science Quarterly* 29 (4): 499-517.

Goddard, Yvonne, Goddard, Roger, and Tschannen-Moran, Megan. 2007. “A Theoretical and Empirical Investigation of Teacher Collaboration for School Improvement and Student Achievement in Public Elementary Schools.” *Teacher College Record* 109 (4): 877-896.

Goodman, P., Rawlin, E., and Schminke, M. 1987. *Understanding Groups in Organizations*. Greenwich, CT: JAI Press.

Grady, Matthew, and Beretavis, S. Natasha. 2010. “Incoporating Student Mobility in Achievment Growth Modeling: A Cross-Classified Multiple Membership Growth Curve Model.” *Multivariate Behavioral Research* 45 (3): 393-419.

Hackman, J. Richard. 1980. Work Redesign and Motivaton. *Professional Psychology,* 11 (3): 445-455.

Hite, Steven. 2001. *Reviewing Quantitative Research to Inform Educational Policy Processes*. Paris: United Nations.

Hord, Shirley. 1997. “Professional Learning Communities: What They Are and Why Are They Important?” *Issues about Change* 6 (1): 8-62.

Hord, Shirley. 1998. *Creating a Professional Learning Community at Cottonwood Creek.* Washington, DC: Southwest Educational Development Lab.

Huffman, Douglas, and Kalnin, Julie. 2002. "Collaborative Inquiry to Make Data-Based Decisions in Schools." *Teaching and Teacher Education,* 19 (2003): 569-580.

Jackson, Sherri. 2015. *Research Methods and Statistics: A Critical Thinking Approach*. Boston: Cengage Learning.

Jones, Lelise, Stall, Greg, and Yarbrough, Debra. 2013. "The Importance of Professional Learning Communities for School Improvement." *Creative Education* 4 (5): 357-361.

Key, Ellen. 2006. "Do They Make a Difference? A Review of the Impact of Critical Friend Groups". Paper presented at the National School Reform Faculty Research Forum. http://www.nsrfharmony.org/research.key.pdf

Kruse, Sharon, and Louis, Karen. 1993. "An Emerging Framework for Analyzing School-Based Professional Community." Paper presented at the American Educational Research Association, Georgia, USA.

Kruse, Sharon, Louis, Karen, and Bryk, Anthony. 1995. *An Emerging Framework for Analyzing School Based Professional Community*. Thousand Oaks, CA: Corwin Press.

Kulik, Carol, and Oldham, Greg. 1987. "Work Design as an Approach to Person-Environment Fit." *Journal of Vocational Behavior* 31 (1987): 278-296.

Lee, Valerie, and Smith, Julia. 1996. "Collective Responsibility for Learning and It's Effects on Gains of Early Secondary School Students." *American Journal of Education* 104 (2): 103-147.

Lewis, Cathrine, and Tsucihida, Ineko. (1997). Planned Educational Change in Japan: The Case of Elementary Science Instruction. *Journal of Educational Policy,* 12 (5): 313-331.

Lieberman, Joanne. 2009. "Reinventing Teacher Professional Norms and Identities: The Role of Lesson Study and Learning Communities." *Professional Development in Education* 35 (1): 83-89.

Little, Judith. 1982. "Norms of Collegiality and Experimentation: Workplace Conditions of School Success." *American Educational Research Journal* 19 (3): 325-340.

Lomos, Catalina, Bosker, Roel, and Hoffman, Roelande. 2012. "Professional Community and Student Acheivement." S. G. Huber & F. Ahlgrimm, *Cooperation in the School*. Munich, Germany: Waxmann.

Lomos, Catalina, Hoffman, Roelande, and Bosker, Roel. 2011a. "Professional Communities and Student Achievement – a Meta-Analysis." *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice* 22 (2): 121-148.

Lomos, Catalina, Hoffman, Roelande, and Bosker, Roel. 2011b. “The Relationship Between Departments as Professional Communities and Student Achievement in Secondary Schools.” *Teaching and Teacher Education* 27 (4): 722-731.

London, Manuel, and London, Marilyn. 1996. “Tight Coupling in High Performing Teams.” *Human Resource Management Review* 6 (1): 1-24.

Louis, Karen. 2006. “Changing the Culture of Schools: Professional Community, Organizational Learning, and Trust.” *Journal of School Leadership* 16 (5): 477-489.

Louis, Karen, and Marks, Helen. 1998. “Does Professional Community Affect the Classroom? Teachers' Work and Student Experience in Restructuring Schools.” *American Journal of Education* 106 (4): 532-575.

McLaughlin, Milbrey. 1979. “Staff Development and School Change.” *Teachers College Record* 80 (1): 60-94.

McLaughlin, Milbrey, and Talbert, Joan. 2006. *Building School Based Teacher Learning Communities: Professional Strategies to Improve Student Achievement*. New York: Teachers College.

Mertens, Donna. 2015. *Research and Evaluation in Education and Psychology:Iintegrating Diversity with Quantitative, Qualitative, and Mixed Methods.* Los Angeles, CA: SAGE Publications.

Mills, G., and Gay, Lorrie. 2015. *Educational Research: Competencies for Analysis and Applications* (11th Edition). London, UK: Pearson.

Mitchell, Cora, and Sackney, Larry. 2001. “Building Capacity for a Learning Community.” *Canadian Journal of Educational Administration and Policy* 1(19): 7-17.

Moller, Stephanie, Mickelson, Roslyn, Stearns, Elizabeth, Banerjee, Neena, and Bottia, Martha 2013. “Collective Pedagogical Teacher Culture and Math Achievement: Difference by Race, Ethnicity, and Socioeconomic Status.” *Sociology of Education* 86 (2): 174-194.

Moolenaar, Nienke, Sleegers, Peter, and Daly, Alan. 2012. “Teaming Up: Linking Collaboration Networks, Collective Efficacy, and Student Achievement.” *Teaching and Teacher Education* 28 (2): 251-262.

Murphy, Joseph. 1990. “Helping Teachers Prepare to Work in Restructured Schools.” *Journal of Teacher Education* 41 (4): 50-56.

Mutch-Jones, Karen, Puttick, Gillian, and Minner, Daphne. 2012. “Lesson Study for Accessible Science: Building Expertise to Improve Practice in Inclusive Science Classrooms.” *Journal of Research in Science Teaching* 49 (8): 1012-1034.

National Council of Teachers of English. 2010. Teacher Learning Communities. *The Council Chronicle* 20(2), 1-4.

Nave, Bill. 2000. *Among Critical Friends: A Study of Critical Friends Groups in Three Maine Schools.* PhD diss. Harvard, Cambridge, MA.

Newmann, Fred, and Wehlage, Gary. 1995. *Successful School Restructuring: A Report to Public Education*. Washington, D.C.: Center on Organization and Restructuring of Schools.

Newmann, Fred, King, Bruce, and Youngs, Peter. 2000. "Development that Addresses Building Capacity." *American Journal of Education* 108 (4): 259 -299.

Pang, Ming, and Ling, Lo. 2012. "Learning Study: Helping Teachers to Use Theory, Develop Professionally, and Produce New Knowledge to be Shared." *Instructional Science* 40 (3): 589-606.

Penuel, William, Riel, Margaret, Krause, Ann, and Frank, Kenneth. 2009. "Analyzing Teachers' Professional Interactions in a School as Social Capital: A Social Network Approach." *Teachers College Record* 111 (1): 124-163.

Raudenbush, Stephen, and Bryk, Anthony. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods* (2nd ed.). Thousand Oaks, CA: SAGE.

Reynolds, Bart A. 2005. "An Analysis of the Evidence Produced by Quantative Research on the Effects of School Discipline and Academic Achievement." PhD diss., Brigham Young University, Provo.

Roberts, Mindy. 2010. "Improving Student Achievement Through Professional Learning Communities" *Educational Administration: Theses, Dissertations, and Student Research.* University of Nebraska, Lincoln, NE: Digital Commons.

Robinson, Naomi, and Leikin, Rosa. 2012. "One Teacher Two Lessons: The Lesson Study Process." *International Journal of Science and Mathematics Education* 10 (1): 139-161.

Runkel, Phillip, Wyant, Spencer, and Bell, Warren. 1975. *Organizational Specialists in a School District: Four Years of Innovation. A Report of the Program on Strategies of Organizational Change.* Washington, DC: Center for Educational Policy.

Sackney, Larry, and Walker, Keith. 2005. *Leadership for Knowledge Communities*. Paper presented at the Commonwealth Council for Educational Administration and Management, Nicosia, Cyprus.

Saunders, William, Goldenberg, Claude, and Gallimore, Ronald. 2009. “Increasing Achievement by Focusing Grade Level Teams on Improving Classroom Learning: A Prospective Quasi Experimental Study of Title I Schools.” *American Educational Research Journal* 46 (4): 1006-1033.

Schmoker, Mike. 2004. “Tipping Point: From Feckless Reform to Substantive Instructional Improvement.” *Phi Delta Kappan* 85 (6):424-432.

Senge, Peter. 1990. “The Leader's New Work: Building Learning Organizations.” *Sloan Management Review* 32 (1): 7-33.

Senge, Peter. 1995. “On Schools as Learning Organizations: A Conversation with Peter Senge.” *Educational Leadership* 52 (7) : 20-25.

Sleegers, Peter, den Brok, Perry, Verbiest, Erick, Moolenaar, Nienko, and Daly, Alan. 2013. “Toward Conceptual Clarity: A Multidimensional, Multilevel Model of Professional Learning Communities in Dutch Elementary Schools.” *The Elementary School Journal* 114 (1): 118-1137.

Smylie, Mark, Wenzel, Stacy, Allensworth, Eliane, Fendt, Carol, Hallman, Sara, Luppescu, Stuart, and Nagaoka, Jenny. 2003. “The Chicago Annenberg Challenge: Successes, Failures, and Lessons for the Future. Consortium on Chicago School Research.” *Improving Chicago's Schools*. Chicago, IL: The Annenberg Foundation.

Stahl, Katherine, Keane, Annette, and Simic, Onjen. 2013. “Translating Policy to Practice: Initiating RTI in Urban Schools.” *Urban Education* 48 (3): 350-379.

Stoll, Louise, and Louis, Karen. 2007. “Professional Learning Communities: Divergence, Depth, and Dilemmas.” In *Professional Learning*, I. Goodson & A. Hargreaves. Berkshire, UK: McGraw Hill Education.

Stollar, Lori. 2014. “Teachers' Perceptions of a Professional Learning Community Model and its Impact on Teaching and Learning.” PhD diss., Widener University, Wilmington.

Sundstrom, Eric, De Meuse, Kenneth, and Futrell, David. 1990. “Work Teams: Applications and Effectiveness.” *American Pyschologist* 45 (2): 120-133.

Taylor, Matthew, Hallam, Pamela, Charlton, Cade, and Wall, Gary. 2014. “Formative Assessment of Collaborative Teams (FACT): Development of a Grade-Level Instructional Team Checklist.” *National Association of Secondary School Principals. NASSP Bulletin* 98 (1): 26-52.

The National Commisson on Excellence in Education 1983. “A Nation at Risk: The Imperative for Educational Reform.” *The Elementary School Journal* 84 (2): 112-130.

Trochim, William. 1998.” Donald T. Campbell and Research Design.” *American Journal of Evaluation* 19 (3): 407-409.

Vescio, Vicky, Ross, Dorene, and Adams, Alyson. 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.

Voelkel, Robert. 2011. “A Case Study of the Relationship Between Collective Efficacy and Professional Learning Communities.” PhD diss., University of California, San Diego.

Wenger, Etienne, Lave, Jean. 1991 *Situated Learning: Legitimate Peripheral Participation*. Cambridge, UK: Cambridge University Press.

Wenger, Etienne, McDermott, Richard, and Snyder, William. 2002. *Cultivating Communities of Practice*. Cambridge, MA: Harvard Business School Press.

Whalan, Francis. 2012. *Collective Responsibility: Redefining what falls between the cracks for school reform.* Boston, MA: Sense Publishers.

Williams, Ellen, Matthews, Joseph, Stewart, Courtney, and Hilton, Sterling. 2007. “The Learning Community Culture Indicator: The Development and Validation of an Instrument to Measure Multidimensional Application of Learning Communities in Schools.” Paper presented at the University Council for Education Administration, Washington, DC.

Wilson, Cate. 2014. “Effective Professional Learning Communities: The Possibilities for Teachers as Agents of Change in Schools.” *British Educational Research Journal* 40 (1): 18-29.

Yoon, Kwang, Duncan, Teresa, Lee, Silvia, Scarloss, Beth, and Shapley, Kathy. 2007. *Reviewing the Evidence on How Teacher Professional Development Affects Student Achievement*. Washington, DC: Regional Educational Laboratory SouthWest.

APPENDIX A: EXTENDED LITERATURE REVIEW

Many reviews of professional community efforts' (*plc*) tie to student achievement limit their focus to those efforts labeled Professional Learning Communities and/or Professional Community (Lomos, Hoffman, & Bosker, 2011a; Yoon, Duncan, Lee, Scarloss, & Shapley, 2007). Using broader search parameters another study found eleven papers qualified for their review (Vescio, Ross, & Adams, 2008). While limited focus creates a tidy universe of studies that can be readily accessed and reviewed, labels can also restrict and mislead. The number of studies examined in such cases is small. Some reform efforts labeled PLCs lack the cooperative characteristics thought to identify learning organizations. Other reform efforts under different labels exhibit many of the essential cooperative characteristics. As in so many other educational topics "professional learning communities (PLCs) and teacher learning teams (LTs) can be traced to many sources" (Gallimore, Ermeling, Saunders, & Goldenberg, 2009, p. 538).

Groundwork for the development of PLCs occurred in the 1980s when concepts of teams and learning organizations were formulated (Goodman, Rawlin, & Schminke, 1987; Hackman, 1980; Kulik & Oldham, 1987; Senge, 1990). Researchers from various fields identified high-work team differentiation, high integration, performance orientation, and self-government as characteristics of high performing groups (London & London, 1996; Sundstrom, De Meuse, & Futrell, 1990). When open classroom teams of the 1970s failed to show expected results (Gamsky, 1970), researchers (McLaughlin, 1979; Runkel, Wyant, & Bell, 1975) explored increasing educational effectiveness and student learning through professional development.

Educational Reform via Professional Development

In 1872 the Japanese government decided to modernize their educational system. To expand their educators' instructional repertoire the *criticism lesson* was introduced into the

teacher training programs (Mutch-Jones, Puttick, & Minner, 2012). In a *criticism lesson* an aspiring teacher presented an object lesson to fellow teachers who evaluated the lesson in terms of content, method, and student response. By 1960 an evolved format was called *lesson study* (Fernandez, Cannon, & Chokshi, 2003). In a *lesson study* session collaborative teams of educators would: (a) select a goal, (b) create a lesson aligned to that goal, (c) present the lesson to the students while colleagues monitored student reactions, (d) gather together afterwards to discuss the presentation of the lesson and the observed student reactions, and (e) make improvements to the lesson plan. In many important ways, Japanese *lesson study* could be considered as *plc*.

The term lesson study shows up in western research studies about 1997 (Lewis & Tsucihida, 1997; Robinson & Leikin, 2012). Long before 1997, the challenge of Japanese competition had reached America's shores. In 1983, Secretary of Education, T.H. Bell, published a report entitled, "A Nation at Risk." Recommendation #7 for teaching was that "master teachers should be involved in designing teacher preparation programs and in supervising teachers during their probationary years" (The National Commission, 1983, p. 9) demonstrating that by 1983 professional development was an acceptable avenue toward educational reform. Staff interaction received attention as researchers discovered that more successful schools exhibited "patterned norms of interaction among staff" (Little, 1982, p. 325).

Collaboration in Educational Reform

Early research on groups of individuals in organizations suggested that decentralization and site-based management were important for effective group development (Murphy, 1990). Seizing upon the insight that "task groups form a link between the individual and the organization" (Gladstein, 1984, p. 499), researchers in 1993 called for professional development

to create communities of educators influencing the direction of school reform efforts at the school level (Kruse & Louis, 1993). Multiple strands of collaborative teacher groups developed. Schools going beyond traditional lab schools by implementing collaborative groups for *both* professional development and educational reform were called *Professional Development Schools* (Darling-Hammond, Bullmaster, & Cobb, 1995). Cosner (2011) describes collaboration in such schools:

> Grade-level teams met weekly during the school day in all three schools and were expected to analyze these assessments and draw upon analysis to monitor student progress and inform instructional planning and decision-making. The university network tasked principals and literacy coordinators to provide leadership for this reform work and expected schools to use three whole-school professional development sessions, scheduled shortly after each of the three cycles of grade-level data-based collaboration, as settings to engage each grade-level team in formal discussions about their data-based collaboration, understandings from, and use of data analysis. (p. 572)

A second collaborative strand was the creation of *Critical Friends Groups* (CFG). In 1994 the Annenberg Institute funded the National School Reform Faculty (NSRF) at Brown University. By 1996 NSRF developed *Critical Friends Groups* along with protocols and associated tools intended to help educators improve their practice (Dunne, Nave, & Lewis, 2000; Key, 2006; Nave, 2000). Dirim (2010) illustrated the use of such protocols:

> Commonly used protocols involve looking at student work in which a teacher brings a sample of student work and presents the work along with a focusing question. Members of the group then take turns describing and hypothesizing about the work while the presenting teacher takes notes. After several rounds of comments, the presenting teacher

> shares what she found useful in the conversation. Then the group debriefs the entire process. Protocols used for peer observation involve two teachers using a predetermined format and focus for observing each other's teaching. Problem-solving protocols open with the presenter asking questions about a specific dilemma. Participants then ask probing questions and discuss the problem among themselves while the presenter takes notes until the discussion is finished, at which point the presenter shares what she heard that was useful or important for her dilemma. All CFG protocols use specific turn-taking rules, and then feedback given is observational, not judgmental. (p. 4)

A third collaborative strand traces its roots back to 1991 when Wenger and Lave studied how a new employee is introduced to a set of best practices in a business context, calling such introductions *situated learning*. By 2002 Wenger placed situated learning under the heading of *community of practice* (Wenger, McDermott, & Snyder, 2002). In an educational context, a *community of practice* seeks to ask and answer the following questions (Wenger, 2006, p. 5):

1. How to organize educational experiences that ground school learning in practice through participation in communities organized around subject matter?
2. How to connect the experience of students to actual practice through peripheral forms of participation in broader communities beyond the walls of the school?
3. How to serve the life-long learning needs of students by organizing communities of practice focused on topics of continuing interest to students beyond the initial schooling period?

Communities of practice with a focus on continuous learning beyond the introductory period (Lieberman, 2009) can become Teacher Learning Communities (TLCs). In Lieberman's TLC, eight to ten teachers at a site agreed to embed formative assessments into their practice.

The teachers met regularly to report on their own progress, to discuss the work of other educators, to consider ways to improve the lesson, and to set goals and objectives for the next round. TLCs needed to be supported by administrators, be teacher driven, be discussing concerns that emerged from the classroom and be motivated by a spirit of inquiry (National Council of Teachers of English, 2010).

In 1990, one year prior to Wenger's work, Peter Senge coined the term *learning organization* (Senge, 1990) to describe an organization that is continually evolving. Senge, writing for business organizations, described the leader's role in a learning organization as that of a teacher. Senge's use of the title *teacher* primed his concept for application to education:

> Leaders are designers, teacher, and stewards. These roles require new skills: The ability to build shared vision, to bring to the surface and challenge prevailing mental models, and to foster more systemic patterns of thinking. In short, leaders in learning organizations are responsible for building organizations where people are continually expanding their capabilities to shape their future, that is, leaders are responsible for learning. (p. 9)

Senge felt that, when applied to education, a learning organization meant redesigning and restructuring the teacher's role (Senge, 1995):

> There is a huge difference between individual capability and collective capacity, and individual learning and collective learning. But this is rarely reflected in the way schools are organized, because education is so highly individualistic. A second dimension of the problem is that educational institutions are designed and structured in a way that reinforces the idea that my job as a teacher is as an individual teaching my kids. (p. 2)

As the prior examples illustrate, many of the reform efforts in the era of collaboration and educational reform were not labeled PLCs, yet they exhibited many of the essential characteristics of *plc*. This era served as a bridge to the era of professional community that led to the era of PLCs.

Professional Community and Educational Reform

Kruse lists the five characteristics of a professional community as: (a) shared values, (b) collaboration, (c) a focus on student learning, (d) de-privatization of practice, and (e) reflective dialog (Kruse & Louis, 1993; Kruse, Louis, & Bryk, 1995). The inclusion criteria for this study used the first three of Kruse's characteristics and combined the fourth and fifth under a heading of *a reflective focus on results to determine best practices*. The following discussion demonstrates that the first three characteristics of professional community echo the concepts of Senge's learning organization and are shared with TLCs as well.

Shared vision and values. Senge envision a integrated organization where leaders developed, among other skills, the new skill of building shared vision throughout the company (Senge, 1990). When school staff construct a shared vision for student learning; develop trusting, interpersonal relationships; and embark upon a program of continuous learning, the staff establishes professional community (Hord, 1997). Shared values provide a foundation for decision making that is open, ethical, collective, and participative (Kruse et al., 1995; Newmann & Wehlage, 1995). The focus of the shared vision should be on all students learning at higher levels (DuFour, 2004; Hord, 1998).

Collective responsibility for student learning. Educational researchers have claimed that professional community aids schools in the development of collective responsibility for student learning (Louis & Marks, 1998), that Senge found so lacking in educational institutions (

1995). In the earliest discussion of professional community, the critical assumption was that the objective was to make a difference in student learning with increased teacher morale and teaching skills as secondary (Stoll & Louis, 2007). Educators who broaden their scope of responsibility soon discover that collective responsibility for student learning is self-reinforcing (Whalan, 2012). One researcher claimed a self-reinforcing cycle developed; as teachers felt efficacious in their instruction, the teachers would expend greater effort helping students learn, which in turn would increase their perception that students are capable learners (Rosenholtz, 1989). In schools where collective responsibility for student learning was high, the research literature claimed that student academic gains were higher than in schools where collective responsibility for students was low (Lee & Smith, 1996; Lomos et al., 2011b; Moolenaar, Sleegers, & Daly, 2012).

Collaborative efforts focused on student learning. Senge noted that the complex challenges faced by a learning organization required collaborative learning among different, but equally qualified individuals (Senge, 1990). Collaboration within a school occurs as teachers share expertise with each other on how to present concepts not yet mastered by the students. An individual teacher may be the expert one moment and the learner the next (Lieberman, 2000). As teachers hone their skills collaboratively to improve student instruction, they can develop a greater trust in one another and in the students (Louis & Marks, 1998). Evidence of collaboration might be found as various course goals are aligned with each other although each course is taught by different instructors (Visschers & Witzers, 2004). Researchers claimed that such collaboration influenced students indirectly with positive yet small effect on student achievement (Goddard, Goddard, & Tschannen-Moran, 2007; Newmann, King, & Youngs, 2000). One way that teachers collaborate is through de-privatization of practice.

De-privatization of practice. Team teaching, classroom observations, and peer coaching are examples of de-privatization of practice. Classroom observation allowed teachers to trade off roles of mentor, advisor, and specialist all with an aim to provide aid and assistance to each other (Kruse & Louis, 1993). De-privatization of practice is the characteristic least likely to be observed in studies of *plc* in schools (Bolam, McMahon, Stoll, Thomas, & Wallace, 2005; Lomos et al., 2011b). More frequently teachers would de-privatize their student formative data in a team or grade-level effort to increase student achievement in a particular subject area. In view of the recent trend of rarity of classroom observations, this characteristic was not among the inclusion criteria for this study, being subsumed by the criteria of collaborative efforts focused on student learning.

Reflective dialog. Reflective dialog requires the educator to view their instruction though both the lens of the teacher and the lens of a researcher. As an example, while working with Japanese educators to implement lessons study into the US curriculum, one researcher (Fernandez et al., 2003) noted:

> We observed the Japanese teachers continually encouraging the American teachers to see themselves as researchers conducting an empirical examination, organized around asking questions about practice and designing classroom experiments to explore these questions. In particular, the Japanese teachers emphasized four critical aspects of good research: the development of meaningful and testable hypotheses, the use of appropriate means for exploring these hypotheses, the reliance on evidence to judge the success of research endeavors, and the interest in generalizing research findings to other applicable contexts. (p. 173)

Attending to reflective dialog reduces the concern noted in a study of 40 districts, which found that community learning time was rarely used to talk about improving instruction (Ermeling & Gallimore, 2013). The concept of the teacher as researcher, with roots going back to the lab school efforts at the University of Chicago, is developed into this study's fourth inclusion criteria as evolved during the PLC era.

Professional Learning Communities (PLCs)

Researchers studied other potential characteristics of *plc* beyond the five originally proposed by Kruse in 1993. When collaboration moved beyond professional development focused on a set of teacher skills, was continuous, and changed the culture of the school, that professional development was called a Professional Learning Community (PLC) (Louis, 2006; Stollar, 2014).

Continuous collective learning/formative data. Changing a school culture required time and multiple iterations of practice (Louis, 2006). In a PLC, teachers would strive to develop not only instructional strategies but also the inclination to continually improve their instructional skills (DuFour, Eaker, & DuFour, 2006). Recent research stressed the importance of focusing on PLCs as a multi-dimensional and multi-layered construct (Sleegers, den Brok, Verbiest, Moolenaar, & Daly, 2013). Feedback between the students and teacher nurtured the continuous learning process. Data from common, formative assessments were proposed as a critical PLC component in 2004 (Richard DuFour, 2004). Feedback as an important factor of learning has been recognized for several decades. A review of feedback literature published up to the year 1992 differentiated between process feedback and result feedback (Kluger & DeNisi, 1996). Process feedback, feedback that relates to task learning or changes in learning strategies, was identified as being strongly associated with high achieving groups (Hattie & Timperley, 2007).

Traditionally teachers provide students results feedback. Results feedback focuses on performance acceptability (such as proficiency level or percent correct). In a PLC, emphasis shifted toward process feedback; for example, a teacher compares their initial approach with alternate strategies suggested in a PLC and gains a heightened sense of self-efficacy by increasing the number of instructional strategies available for employment during instruction. A reflective focus on results to determine best practices was the fourth inclusion criteria for this study. Beyond these critical four criteria, two additional *plc* characteristics were noted in the literature, although they were not required for inclusion in the census for this study.

Shared leadership. Hord contended that adding shared leadership to Kruse's characteristics of a professional community created a PLC (DuFour & Eaker, 1998; Hord, 1998; Hord, 1997). Gains hinted at in earlier research (Darling-Hammond, 1994) required some decentralization of decision making. A PLC expected those closest to the instruction to control the allocation of resources directly influencing instruction (Sackney & Walker, 2005). Mitchell argued that a PLC requires that leadership be spread throughout the school in a many different ways (Mitchell & Sackney, 2001).

High trust/supportive structures. As teachers honed their skills collaboratively to improve student instruction, they developed a greater trust in one another and in the students (Louis & Marks, 1998). Trust must also exist between teachers and administrators sufficient for teachers to feel comfortable asking for help and exposing weaknesses (Byrk & Schnieder, 2002). Confidence to voice a lack of knowledge signaled to researchers that the group was indeed a learning community (Richmond & Manokore, 2010). A PLC creates additional levels of trust and allows for advanced levels of community (Huffman & Kalnin, 2002). In order to foster high levels of trust, appropriate temporal and social structures were needed to encourage success of

the PLC (Stoll & Louis, 2007). Proximity and adequate time for discussion were listed as critical components for effective PLCs (Bolam et al., 2005). For example, one study on social capital and professional community networks found the greatest variance in student achievement hinged on the proximity of veteran and novice teachers in learning teams (Penuel, Riel, Krause, & Frank, 2009).

Essential Characteristics of *plc*

A *plc* is a complex phenomenon (Wilson, 2014), and debate rightly continues over what constitutes its critical components thereby avoiding the "cold comfort of final definition" (Clegg, Kornberger, & Rhodes, 2005, p. 149) and allowing deeper insight into the restructuring of teacher roles. Commenting on the eight characteristics of an effective PLC presented to United Kingdom's Department of Education and Skills in 2005, Wilson notes that the most common characteristic of a PLC was shared mission and values
(Bolam et al., 2005; Wilson, 2014). The next three most frequently noted characteristics were collective responsibility for pupils' learning, collaboration focused on student learning, and reflective professional inquiry. Wilson argues that while other characteristics (high trust, shared leadership, etc.) may be critical for effective PLCs these characteristics are less frequently found (Bolam et al., 2005). In the *Method* section, Appendix B, I outline the inclusion criteria for a document to be included in the census of articles on *plc* and student achievement. The first three criteria are essentially the same ones listed by Wilson with the fourth being modified to reflect DuFour's contention that PLCs must be data driven with formative data driving intervention and remediation strategies (Richard DuFour, 2004; Wilson, 2014).

APPENDIX B: DETAILED METHODS

Many reviews of *plc* and student academic achievement limit their focus to those efforts explicitly labeled Professional Learning Communities (upper case PLC) and/or Professional Community (Lomos et al., 2011a; Yoon et al., 2007). A broader set of search parameters used in another study, allowing *critical friends groups* and *communities of practice*, found only 11 papers qualified for review (Vescio et al., 2008) and eight of the 11 dealt with student academic gains. While such a limited focus creates a tidy sample of studies to be readily accessed and reviewed, labels can also be overly restrictive and potentially misleading. For example, some reform efforts labeled PLC lack the cooperative characteristics claimed by experts to identify bone fide PLCs. Other reform efforts, under different labels, exhibit many of the essential cooperative characteristics and perhaps additional characteristics beyond those in a strictly identified PLC. As in so many other educational topics, "professional learning communities (PLCs) and teacher learning teams (LTs) can be traced to many sources" (Gallimore et al., 2009, p. 538) and be defined in various ways.

This analysis of *plc* research was based on a census of quantitative, primary research articles in English-language scholarly journals published between January 1, 1980, and January 1, 2015. The decision tree for inclusion of a given article is shown in Figure 1.

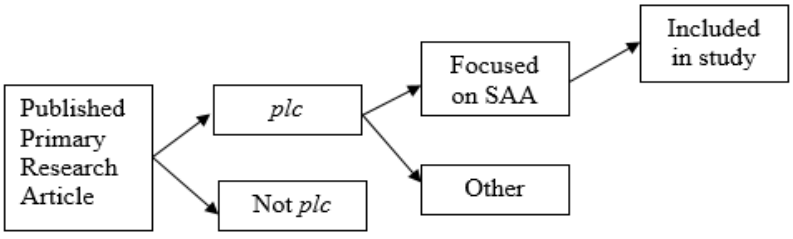

Figure 1. Planning/evaluation cycle.

The analysis of each article in this census first determined the design type of each research document, that reported student academic achievement findings in the presence of *plc*. Next, this study analyzed the likelihood that the *plc* created the student gains based on typical design expectations and limitations regarding causal claims (Jackson, 2015; Mills & Gay, 2015).

The two inclusion criteria for research articles in the census for this study included: (a) published quantitative, primary results concerning *plc* and student academic achievement; (b) studies that exhibited, at least, the four following *plc* characteristics (Bolam et al., 2005; Richard DuFour, 2004; Wilson, 2014):

1. shared vision and value,
2. collective responsibility for and focus on student learning,
3. collaborative teaming and learning efforts, and
4. a reflective focus on results to determine best practices.

This census included all efforts identified as *plc* based on the above inclusion criteria, even those not specifically identified as professional learning communities. Rather than a handful of studies that qualified (Yoon et al., 2007), this census resulted in 57 research studies. The decision to use a broad definition of *plc* (along with the lower case acronym for such efforts) was purposeful and facilitated the following advantage over previous efforts: (a) no effective criticism can be leveled that the outcome of the analysis was predetermined or manipulated towards any particular objective by excluding favorable or unfavorable studies form the census through restrictive inclusion criteria; (b) no data from the more constrained studies has been lost to this analysis; and (c) there was decreased chance of parochial or cultural misinterpretation of the findings since there is no set of trademarked buzzwords used to determine inclusion.

The focus on quantitative results focused on student academic achievement is reasonable because so many of those promoting *plc* argue for implementation precisely because it will increase student achievement. This is not to say that there are no other claimed benefits to *plc,* but increased student learning is a major focus in much of the literature.

Data Collection Procedures

Over 200 potentially qualifying articles were reviewed from searches on the Internet, EBSCO database, Web of Science database, and ERIC. The online searches used the following keywords in various combinations: Achievement, Communities of Practice, Critical Friends Group, Gains, Learning Community, Lesson Study, Multi-Tier Systems of Support, PLC, Professional Community, Response to Intervention, Teacher Learning Community, and/or Teacher Networks. Studies not looking at quantitative measures of student academic achievement as a dependent variable or containing no primary student achievement data were eliminated. The remaining articles were examined to see if collaborative efforts described qualified for inclusion as *plc* by exhibiting at least the four characteristics in the inclusion criteria, resulting in 57 articles in the census.

Data Analysis

Each of the qualifying articles was evaluated using an adaptation of Reynolds' design and implementation matrix, itself based on Hite's checklist for reviewing research documents (Hite, 2001; Reynolds, 2005). Figure 2 illustrates the decision tree for the review of the published research articles. Study design components were marked as either being present in the article (Yes) or not found (No). The quality of the implementation of the corresponding components of design was marked with Low, Medium or High (1, 2, or 3) based on expectations put forth in

typical educational research design texts (Creswell, 2014; Gay, Airasian, & Mills, 2012; Mills & Gay, 2015). Note that all scores are based only on published details in the collected articles.

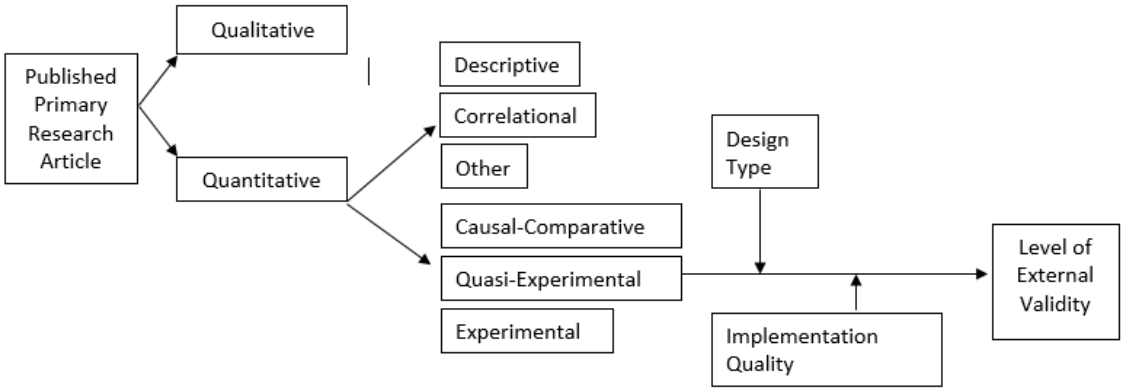

Figure 2. Decision tree for review of published primary research articles.

Table 5 provides the Design and Evaluation Matrix. The *Design Elements* section of the matrix consisted of a check mark for a YES response and a blank box for a NO response. The *Quality of Implementation* section of matrix used the ordinal numbers 1, 2, and 3 to represent Low, Medium, and High ratings. A high rating indicated high likelihood of quality, and a low rating indicated low (or lack of) quality indicators. The rubric used to assign ratings of High, Medium or Low for each section of the Design and Implementation Evaluation Matrix is provided in Table 6.

Validity

Since validity is one of the major methodological focal points of this study, a review of the basic constructs of validity is needed. Validity, as presented here, relates specifically to how contemporary quantitative researchers construct this term. A discussion of validity in ontologically qualitative terms is not germane to this discussion.

Table 5

Design and Implementation Evaluation Matrix

Design Element	**Yes/No**
Purpose and problem(s)	Purpose Found and noted? Research Problems found and noted? Variables specified? Context of research clarified?
Research method and methodology	Methods presented and discussed? Methodology described (descriptive, correlational, causal-comparative, quasi-experimental, or experimental)?
Quality of Implementation	**Low / Medium / High** **1 / 2 / 3**
General Procedures	Procedures sufficiently detailed? Procedures reasonable for study?
Sampling	Population defined? Randomly selected?
Instrumentation	Instrumentation reliable/valid? Instrumentation appropriate for context? Training of researchers using instrument adequate?
Data Analysis	Analysis suitable for methodology? Statistical information complete? Data appropriate to context? Statistical significance discussed?
Results	Each research problem addressed? All information in the original?
Conclusions	Internal validity? External validity? Conclusions warranted?

Internal and External Validity

Threats to internal validity are usually classified under headings such as ambiguous temporal precedence, selection bias, history, maturation, pre-test/post-test, instrumentation change, mortality, regression to the mean, and diffusion effects. Threats to external validity are usually classified under headings such as reactivity and experimental effects.

Table 6

Rubric for Design and Implementation Matrix Elements

Matrix Elements	Evaluation Criteria	Low Rating	Medium Rating	High Rating
General Procedures	Procedures sufficiently detailed	Important details are missing or details are scant.	Educator would substitute for some missing steps to replicate study.	Educator would be reasonably able to replicate the study.
	Procedures reasonable for the study	Procedures are appropriate for one of the three (listed in High Rating).	Procedures are appropriate for two out of the three (listed in High Rating).	Procedures are appropriate for the context, grade level(s) and research question.
Sampling	Population is defined	The intended target population must be assumed by the reader.	The target population can be divined by studying the author's claims.	The intended target population is clearly delineated and fully described
	Sample selected randomly	The sample was convenient or clusters self-selected themselves.	Every cluster of members had an equal chance to be included in the sample.	Every member of population had an equal, independent chance to be included.
Instrumentation	Reliable and valid	Assessment was developed by the authors and reliability unsubstantiated.	Assessment was checked for reliability. Assumption of validity made	Assessment is widely recognized as both statistically reliable and valid.
	Reasonable for the context	Requires high levels of guessing. Not designed to measure desired parameters.	Appropriate for student grade level. Approximates measuring parameters.	Appropriate for student skill level. Measures desired parameters accurately.

Matrix Elements	Evaluation Criteria	Low Rating	Medium Rating	High Rating
	Training of researchers adequate	No mention of training for evaluators.	Evaluators experienced using instrument. Reliability appears good.	Evaluators trained on instrument. Checked for inter-rater reliability.
Data Analysis	Analysis suitable for methodology	Misapplication of statistical test for the study design.	Statistical test is appropriate for either the study design or data collected.	Statistical test is appropriate for both the study design and data collected.
	Statistical information complete	Only mean, n-size, and significance or effect size are reported.	Most of the critical statistical information is reported.	Reports mean, SE, n-size of each group, normality analyses, and effect sizes.
	Data used are appropriate for design	Statistics based on data inconsistent with design type.	Most of the statistics fit design type and research problem.	Statistics presented are internally consistent with research problem.
	Statistical significance considered	No test of significance reported.	Appropriate test statistic for design type but low bar for H0 rejection.	Appropriate selection of Chi squared, F, or T statistic for design type.
Results	Purpose of each research question addressed	New research question emerges from the results.	Research question discussions are intermixed.	Each research question discussed individually.
	All information available in original	Focus is mainly on results with perfunctory discussion of process and data.	Only detailed process description available in document or from authors.	Original data and process description available in document or from authors.

Matrix Elements	Evaluation Criteria	Low Rating	Medium Rating	High Rating
Conclusions and Generalizability	Internal validity high	Fails to address several potential confounding influences (listed in High Rating).	Addresses 2 or 3 of the confounding influences (listed in High Rating).	Accounts for non-randomized sampling, Attrition, Historicity, and Teacher Intervention confounding factors.
	External validity high	Low internal validity or non-causal design type.	Medium internal validity and causal design type.	High internal validity, causal design type, and insignificant cultural limitations.
	Conclusions adequately warranted	Conclusions not supported by data analysis.	Each conclusion aligns to design type and most supported by data analysis.	Each conclusion aligns to design type and supported by rigorous data analysis.

**NOTE: Given the ratings are ordinal in nature, comparison between study's aggregated points is not meaningful. Ratings are useful for giving focus to areas of strength in a study or when comparing two studies on the same evaluative category.

Ambiguous temporal precedence is a threat when the targeted outcome cannot be shown to change only after the administration of the treatment. In some research studies the levels of teacher personal efficacy are found to be higher in schools with high student achievement. Is the high student achievement a result of high levels of teacher personal efficacy, or do teachers feel more efficacious in schools where high achievement is widespread? Studies with high internal validity can control for ambiguous temporal precedence by including schools with a fairly recent history of high achievement along with schools with a tradition of high achievement.

Selection bias occurs when randomization of students included in the sample is not achieved. Volunteer sampling greatly increases the risk of bias in selection. Selecting control groups of students in the same manner as the treatment group and with essentially the same characteristics help control for this threat.

Historical effects threaten validity when a unique set of initial factors change dramatically over the course of the research. Longitudinal studies covering a period of dramatic changes in curriculum or social custom are subject to such a threat.

Maturation threats to validity occur when the natural development of student skill sets is driven by changes to the individual over time. Complex changes in the brain physiology occur as children age and develop greater ability to think in abstract terms. Longitudinal studies are particularly vulnerable to this threat unless a similar control group is followed as well.

Testing effects threaten validity when exposure to a pre-test heightens a participant's sensitivity to or interest in a particular skill. A study that surveyed teachers in a faculty meeting to determine the amount of *de-privatization of practice* in the school might classify the school as *low in professional community* while, following the faculty meeting, the teachers implement

increased collaboration after having once been exposed to the concept. A control group helps evaluate the importance of this threat.

Instrumentation changes threaten validity when the test form is changed. Studies running more than one year often find their students testing at different grade levels. Statistical compensation should be used to equate the two test forms before calculating any student achievement gains.

Mortality threats occur when selected students leave the sample. High non-response rates and high levels of participant turnover are both warning flags that this threat needs to be considered. Checking the descriptive statistics of both the pre- and post-mortality group can flag any patterns of potential bias. Are students of lower social economic status more likely to change schools than students from higher income homes?

When the research design was quasi-experimental, quality of implementation was high, and comparative data was used to sufficiently combat all potential threats to validity, the article received a High (3) for both (internal and external) validity questions in the Conclusions box. When threats to validity were not sufficiently counteracted when using a causal comparative study design, the research received a Low (1) or a Medium (2) for both validity questions but might, if conclusions were carefully drawn by the researchers, receive a High (3) on Conclusions being warranted.

APPENDIX C: AUTHOR GUIDELINES (Chicago Style)

CITATIONS IN TEXT

All sources listed in the reference list must be cited at appropriate points in the text by the author's last name, publication year, and pagination where appropriate, as indicated. When author's name is in text: `Rader (1975)`. When author's name is not in text: `(Wills 1976)`. When citing pagination (for a quote): `(Bell 1967, 62)`. With dual authorship, give both names; for three or more, use "et al." For institutional authorship, identify from the beginning of the complete citation: `(U.S. Bureau of the Census 1963, 117)`. When more than one reference to an author was published in one year, distinguish them by use of letters (a, b) attached to the publication year: `(1956a)`. Enclose a series of references (in alphabetical order) within one pair of parentheses, separated by semicolons: `(Bowles and Gintis 1976; Coleman 1973a; Kaiser 1964)`.

REFERENCE FORMAT

List all items alphabetically by author and, within author, by publication year on a separate page titled "References." Do not use APA style. Examples of common references follow:

BOOKS AND CHAPTERS IN BOOKS

Bowles, Samuel, and Herbert Gintis. 1976. *Schooling in Capitalist America.* New York: Basic Books.

Dorsey, John, ed. 1980. *On Mencken*. New York: Knopf.

Kaiser, Ernest. 1964. "The Literature of Harlem." In *Harlem: A Community in Transition,* J. H. Clarke. New York: Citadel.

JOURNALS AND PERIODICALS

Coleman, James S. 1973a. "Loss of Power." *American Sociological Review* 38 (February): 1-17.

Karp, S. 1991. "Is All Black and All Male Right?" *Z Magazine*, June.

Smith, Charles. "The Right and Wrong Way to Teach." *New York Times*, March 8.

DISSERTATION

Trent, James Williams. 1964. "The Development of Intellectual Disposition within Catholic Colleges." PhD diss., University of California, Berkeley.

VARIOUS OTHERS (INCLUDING INTERNET SITES)

Farber, Donna. 1998. "Content Analysis of Teaching Materials." Working Paper 35, College of Education, University of Michigan.

Franklin, Quentin. 1983. "Food: Its Effect on Student Performance." Paper presented at the annual meeting of the American Educational Research Association, Montreal, April 22.

Hall, Sherrie. 2002. "Inkster Schools Face State Takeover." *Detroit News*, April 10, http://www.detnews.com/2002/schools/0204/10.

Radclife, H. 1981. "Literacy and Sandwiches: Significant Correlation?" Unpublished manuscript, Harvard University, Cambridge, MA.

Note: link to preparation of tables guidelines:
http://www.press.uchicago.edu/infoServices/prep-table.html

APPENDIX D: DISSERTATION REFERENCES

Boaler, J., & Staples, M. (2008). Creating mathematical gutures through an equitable teaching approach: The case of Railside school. *Teachers College Record*, *110*(3), 608-645.

Bolam, R., McMahon, A., Stoll, L., Thomas, S., & Wallace, M. (2005). *Creating and sustaining effective professional learning communities*. Bristol, UK: University of Bristol.

Byrk, A., & Schnieder, B. (2002). *Trust in schools*. New York, NY: Russell Sage.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago, IL: Rand McNally.

Campbell, D T., & Stanley, J. C. (1984). *Experimental and quasi-experimental designs for research*. Boston, MA: Houghton Mifflin.

Clegg, S., Kornberger, M., & Rhodes, C. (2005). Learning, becoming, organizing. *Organization*, *12*(2), 147-167.

Cosner, S. (2011). Teacher learning: Instructional considerations and principal communication: lessons from a longitudinal study of collaborative data use by teachers. *Educational Management Administration & Leadership, 39*(5), 568-589. doi:10.1177/1741143211408453

Creswell, J. (2014). *Educational research* (4th ed.). Essex, UK: Pearson Education Limited.

Daly, A. J., Moolenaar, N. M., Der-Martirosian, C., Canrinus, E., & Chrispeels, J. (2011). A capital investment: The effects of teacher human and social capital on student achievement in improving schools. In *International Congress on School Effectiveness and Improvement*, Lymessoll, Cyprus.

Darling-Hammond, L. (1994). Performance-based assessment and educational equity. *Harvard Educational Review*, *64*(1), 5-31.

Darling-Hammond, L., Bullmaster, M., & Cobb, V. (1995). Rethinking professional leadership through Professional Development Schools. *The Elementary School Journal*, *96*(1), 87-106.

Desimone, L. M., Smith, T. M., & Phillips, K. J. R. (2013). Linking student achievement growth to professional development participation and changes in instruction: A longitudinal study of elementary students and teachers in Title I schools. *Teachers College Record*, *115*(5), 1-46.

Dirim, N. (2010). Fostering development of expertise through Critical Friends Group (CFG). Retrieved from http://workbloglearning.wordpress.com/2010/12/10/162/

DuFour, R. (2004). What is a "Professional Learning Community"? *Educational Leadership*, *61*(8), 6-11.

DuFour, R., & Eaker, R. (1998). *Professional Learning Communities at work: Best practices for enhancing student achievement.* Bloomington, IN: Solution Tree.

DuFour, R., Eaker, R., & DuFour, R. (2006). *Learning by doing: A handbook for professional learning communities at work*. Bloomington, IN: Solution Tree.

Dunne, F., Nave, B., & Lewis, A. (2000). Critical Friends Groups: Teachers helping teachers improve student learning. *Phi Delta Kappan*, *28*(4), 31-37: Phi Delta Kappa International.

Ermeling, B., & Gallimore, R. (2013). Learning to be a community: Schools need adaptable models to create successful programs. *JSD: The Learning Forward Journal*, *34*(2), 42-45.

Fernandez, C., Cannon, J., & Chokshi, S. (2003). A US-Japan lesson study collaboration reveals critical lenses for examining practice. *Teaching and Teacher Education*, *19*(2), 171-185. doi: 10.1016/s0742-051x(02)00102-6

Fielding, A., & Goldstein, H. (2006). Cross-classified and multiple membership structures in multilevel models: An introduction and review. In *Department for Education and Skills* (Ed.), Research (Research Report 791). Birmingham, UK: University of Birmingham.

Fullan, M. (2002). The change leader. *Educational Leadership*, *59*(8), 16-20.

Gallimore, R., Ermeling, B., Saunders, W., & Goldenberg, C. (2009). Moving the learning of teaching closer to practice: Teacher education implications of school-based inquiry teams. *The Elementary School Journal, 109*(5), 537-553.

Gamsky, N. (1970). Team Teaching, student achievement and attitudes. *The Journal of Experimental Education*, *39*(1), 32-45.

Gay, L., Airasian, P., & Mills, G. (2012). *Educational research* (10th ed.) New York, NY: Pearson Education.

Gladstein, D. L. (1984). Groups in context: A model of task group effectiveness. *Administrative Science Quarterly*, *29*(4), 499-517.

Goddard, Y. L., Goddard, R. D., & Tschannen-Moran, M. (2007). A theoretical and empirical Iivestigation of teacher collaboration for school improvement and student achievement in public elementary schools. *Teacher College Record*, *109*(4), 877-896.

Goodman, P., Rawlin, E., & Schminke, M. (1987). *Understanding groups in organizations*. Greenwich, CT: JAI Press.

Grady, M. W., & Beretvas, S. N. (2010). Incorporating student mobility in achievement growth modeling: A cross-classified multiple membership growth curve model. *Multivariate Behavioral Research*, *45*(3), 393-419.

Hackman, J. R. (1980). Work redesign and motivaton. *Professional Psychology*, *11*(3), 445-455.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81-112.

Hite, S. (2001). *Reviewing quantitative research to inform educational policy processes*. Paris: United Nations.

Hord, S. (1997). Professional Learning Communities: What they are and why are they important? *Issues about Change*, *6*(1), 8-62.

Hord, S. (1998). *Creating a Professional Learning Community at Cottonwood Creek*. Washington, DC: Southwest Educational Development Lab.

Huffman, D., & Kalnin, J. (2002). Collaborative inquiry to make data-based decisions in schools. *Teaching and Teacher Education*, *19*(6), 569-580. doi: 10.1016/S0742-051X(03)00054-4

Jackson, S. (2015). *Research methods and statistics: A critical thinking approach*. Boston, MA: Cengage Learning.

Jones, L., Stall, G., & Yarbrough, D. (2013). The importance of Professional Learning Communities for school improvement. *Creative Education*, *4*(5), 357-361.

Key, E. (2006). Do they make a difference? A review of the impact of Critical Friend Groups. Paper presented at the National School Reform Faculty Research Forum. http://www.nsrfharmony.org/research.key.pdf

Kluger, A., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. Pyschological Bulletin, 119(2), 254-284.

Kruse, S., & Louis, K. (1993). An emerging framework for analyzing school-based professional community. Paper presented at *The American Educational Research Association*, Georgia, USA.

Kruse, S., Louis, K., & Bryk, A. (1995). *An emerging framework for analyzing school based professional community*. Thousand Oaks, CA: Corwin Press.

Kulik, C. T., & Oldham, G. R. (1987). Work design as an approach to person-environment fit. *Journal of Vocational Behavior*, *31*(3), 278-296.

Lee, V., & Smith, J. (1996). Collective responsibility for learning and it's effects on gains of early secondary school students. *American Journal of Education*, *104*(2), 103-147.

Lewis, C., & Tsucihida, I. (1997). Planned educational change in Japan: The case of elementary science instruction. *Journal of Educational Policy*, *12*(5), 313-331.

Lieberman, A. (2000). Networks as learning communities: Shaping the future of teacher development. *Journal of Teacher Education*, *51*(3), 221-227. doi: 10.1177/0022487100051003010

Lieberman, J. (2009). Reinventing teacher professional norms and identities: The role of Lesson Study and learning communities. *Professional Development in Education*, *35*(1), 83-89.

Little, J. (1982). Norms of collegiality and experimentation: Workplace conditions of school success. *American Educational Research Journal*, *19*(3), 325-340.

Lomos, C., Bosker, R. J., & Hoffman, R. (2012). *Professional community and student achievement* (Unpublished doctoral dissertation). Rijksuniversiteit, Groningen, Netherlands.

Lomos, C., Hoffman, R., & Bosker, R. (2011a). Professional communities and student achievement – A meta-analysis. *School Effectiveness and School Improvement: An International Journal of Research*, *Policy and Practice*, *22*(2), 121-148.

Lomos, C., Hoffman, R., & Bosker, R.l (2011b). The relationship between departments as professional communities and student achievement in secondary schools. *Teaching and Teacher Education*, *27*(4), 722-731.

London, M., & London, M. (1996). Tight coupling in high performing teams. *Human Resource Management Review*, *6*(1), 1-24.

Louis, K. (2006). Changing the culture of schools: Professional community, organizational learning, and trust. *Journal of School Leadership*, 16(5), 477-489.

Louis, K., & Marks, H. (1998). Does professional community affect the classroom? Teachers' work and student experience in restructuring schools. *American Journal of Education*, *106*(4), 532-575. http://www.jstor.org/stable/1085627

McLaughlin, M. (1979). Staff development and school change. *Teachers College Record*, *80*(1), 69-94.

McLaughlin, M., & Talbert, J. (2006). *Building school based teacher learning communities: Professional strategies to improve student achievement*. New York, NY: Teachers College Press.

Mertens, D. (2015). *Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed methods* (4 ed.). Thousand Oaks, CA: SAGE Publications, Inc.

Mills, G., & Gay, L. (2015). *Educational research: Competencies for analysis and applications* (11th Edition). London, UK: Pearson.

Mitchell, C., & Sackney, L. (2001). Building capacity for a learning community. *Canadian Journal of Educational Administration and Policy*, *1*(19), 7-17.

Moller, S., Mickelson, R., Stearns, E., Banerjee, N., & Bottia, M. C. (2013). Collective pedagogical teacher culture and math achievement: Difference by race, ethnicity, and socioeconomic status. *Sociology of Education*, *86*(2), 174-194.

Moolenaar, N. M., Sleegers, P. J. C., & Daly, A. J. (2012). Teaming up: Linking collaboration networks, collective efficacy, and student achievement. *Teaching and Teacher Education*, *28*(2), 251-262. doi: 10.1016/j.tate.2011.10.001

Murphy, J. (1990). Helping teachers prepare to work in restructured schools. *Journal of Teacher Education*, *41*(4), 50-56.

Mutch-Jones, K., Puttick, G., & Minner, D. (2012). Lesson study for accessible science: Building expertise to improve practice in inclusive science classrooms. *Journal of Research in Science Teaching, 49*(8), 1012-1034. doi: 10.1002/tea.21034

National Council of Teachers of English. (2010). Teacher learning Communities. *The Council Chronicle*, *20*(2), 1-4.

Nave, B. (2000). *Among critical friends: A study of Critical Friends Groups in three Maine schools* (Unpublished doctoral disseratation). Harvard, Cambridge, MA.

Newmann, F., & Wehlage, G. (1995). *Successful school restructuring: A report to public education.* Washington, D.C.: Center on Organization and Restructuring of Schools.

Newmann, F., King, B., & Youngs, P. (2000). Development that addresses building capacity. *American Journal of Education*, *108*(4), 259-299.

Pang, M. F., & Ling, L. M. (2012). Learning study: Helping teachers to use theory, develop professionally, and produce new knowledge to be shared. *Instructional Science*, *40*(3), 589-606. doi: 10.1007/s11251-011-9191-4

Penuel, W., Riel, M., Krause, A., & Frank, K. (2009). Analyzing teachers' professional interactions in a school as social capital: A social network approach. *Teachers College Record*, *111*(1), 124-163.

Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: SAGE.

Reynolds, B. (2005). *An analysis of the evidence produced by quantative research on the effects of school discipline and academic achievement* (Doctoral dissertation). Brigham Young University, Provo, UT.

Richmond, G., & Manokore, V. (2010). Identifying elements critical for functional and sustainable Professional Learning Communities. *Science Education*, *95*(3), 543-570.

Roberts, M. (2010). Improving student achievement through Professional Learning Communities. *Educational Administration: Theses, Dissertations, and Student Research*. University of Nebraska, Lincoln, NE: Digital Commons.

Robinson, N., & Leikin, R. (2012). One teacher, two lessons: The lesson study process. *International Journal of Science and Mathematics Education*, *10*(1), 139-161. doi: 10.1007/s10763-011-9282-3

Rosenholtz, S. (1989). Workplace conditions that affect teacher quality and commitment: Implications for teacher induction programs. *The Elementary School Journal*, *89*(4), 420-439.

Runkel, P., Wyant, S., & Bell, W. (1975). *Organizational specialists in a school district: Four years of innovation. A report of the Program on Strategies of Organizational Change*. Washington, DC: Center for Educational Policy.

Sackney, L., & Walker, K. (2005). Leadership for knowledge communities. Paper presented at *The Commonwealth Council for Educational Administration and Management*, Nicosia, Cyprus.

Saunders, W., Goldenberg, C., & Gallimore, R. (2009). Increasing achievement by focusing grade level teams on improving classroom learning: A prospective quasi-experimental study of Title I schools. *American Educational Research Journal*, *46*(4), 1006-1033. doi: 10.3102/0002831209333185

Schmoker, M. (2004). Tipping point: From feckless reform to substantive instructional improvement. *Phi Delta Kappan*, *85*(6), 424-432.

Senge, P. (1990). The leader's new work: Building learning organizations. *Sloan Management Review*, *32*(1), 7-33.

Senge, P. (1995). On schools as learning organizations: A conversation with Peter Senge. *Educational Leadership*, *52*(7), 20-25.

Sleegers, P. J. C., den Brok, P., Verbiest, E., Moolenaar, N. M., & Daly, A. (2013). Toward conceptual clarity: A multidimensional, multilevel model of professional learning communities in Dutch elementary schools. *The Elementary School Journal*, *114*(1), 118-137.

Smylie, M., Wenzel, S., Allensworth, E., Fendt, C., Hallman, S., Luppescu, S., & Nagaoka, J. (2003). The Chicago Annenberg Challenge: Successes, failures, and lessons for the future. In *Consortium of Chicago Schools Research, Improving Chicago's schools*. Chicago, IL: The Annenberg Foundation.

Stahl, K. A. D., Keane, A. E., & Simic, O. (2013). Translating policy to practice: Initiating RTI in urban schools. *Urban Education*, *48*(3), 350-379.

Stoll, L., & Louis, K. (2007). Professional Learning Communities: Divergence, depth, and dilemmas. In I. Goodson & A. Hargreaves (Eds.), *Professional Learning* (pp. 1-13). Berkshire, UK: McGraw Hill Education. Retrieved from http://web.a.ebscohost.com.erl.lib.byu.edu/ehost/ebookviewer/ebook/bmxlYmtfXzIzNDI1Ml9fQU41?sid=469809b6-e621-4a0c-a4e5-a609834a7583@sessionmgr4005&vid=1&format=EB&lpid=lp_199&rid=0.

Stollar, L. (2014). *Teachers' perception of a Professional Learning Community model and its impact on teaching and learning* (Doctoral dissertation). Widener University, Ann Arbor, MI. (3580319)

Sundstrom, E., De Meuse, K., & Futrell, D. (1990). Work teams: Applications and effectiveness. *American Pyschologist*, *45*(2), 120-133.

Taylor, M. J., Hallam, P. R., Charlton, C. T., & Wall, D. G. (2014). Formative Assessment of Collaborative Teams (FACT): Development of a grade-level instructional team checklist. National Association of Secondary School Principals. *NASSP Bulletin*, *98*(1), 26-52.

The National Commission on Excellence in Education (1983). *A Nation at risk: The imperative for educational reform.* Washington, DC: US Department of Education Retrieved from http://www2.ed.gov/pubs/NatAtRisk/index.html.

Trochim, W. (1998). Donald T. Campbell and research design. *American Journal of Evaluation*, *19*(3), 407-409.

Vescio, V., Ross, D., & 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. doi: 10.1016/j.tate.2007.01.004

Visschers, A., & Witzers, B. (2004). Subject departments as professional communities? *British Educational Research Journal*, *30*(6), 785-799.

Voelkel, R. (2011). *A case study of the relationship between collective efficacy and Professional Learning Communities* (Unpublished doctoral dissertation. University of California San Diego, CA.

Wenger, E., & Lave, J. (1991) *Situated learning: Legitimate peripheral participation*. Cambridge, UK: Cambridge University Press.

Wenger, E., McDermott, R., & Snyder, W. (2002). *Cultivating Communities of Practice*. Cambridge, MA: Harvard Business School Press.

Wenger, E. (2006). *Communities of Practice: A brief introduction*. available from http://www. ewenger.com/theory/communities_of_practice_intro. htm.

Whalan, F. (2012). *Collective responsibility: Redefining what falls between the cracks for school reform* (1st ed.). Boston, MA: Sense Publishers.

Whalstrom, K., & Louis, K. (2008). How teachers experience principal leadership: The roles of professional community, trust, efficacy, and shared responsibility. *Education Administration Quarterly*, *44*(4), 458-495.

Wilson, C. (2014). Effective Professional Learning Communities: The possibilities for teachers as agents of change in schools. *British Educational Research Journal*, *40*(1), 18-29.

Yoon, K., Duncan, T., Lee, S., Scarloss, B., & Shapley, K. (2007). *Reviewing the evidence on how teacher professional development affects student achievement*. Washington, DC: Regional Educational Laboratory SouthWest.