

Facilitating collaborative reflective inquiry amongst teachers: what do we currently know?

Professor Chris Brown¹, Dr. Cindy Poortman², Dr. Helen Gray³, Dr. Jana Groß Ophoff⁴ and Mr. Matthew Wharf⁵

¹ Durham University, England

² University of Twente, Netherlands

³ Durham University, England

⁴ University of Tübingen, Germany

⁵ Plymouth University, England

Corresponding author:

Professor Chris Brown

School of Education

Durham University

Leazes Road

Durham

DH1 1TA

United Kingdom

tel: +00 44 (0) 191 3348195

chris.brown@durham.ac.uk

Abstract

Collaborative teacher learning is thought to improve teaching practice and student outcomes. Key to such learning is Reflective Professional Inquiry (RPI); seen as vital if practitioners are to engage effectively with new knowledge and ideas. Yet RPI is under-conceptualised and little is known about how to facilitate effective RPI. With this study we engage in a meta-narrative literature review, covering a range of disciplines (including education, medicine, and psychology), in an attempt to fill these knowledge gaps. Findings indicate that there are no existing interventions designed to foster RPI that have been rigorously evaluated. Consequently, there are no approaches that could be employed as part of collaborative teacher learning, with the expectation that practice or student outcomes will subsequently improve.

Key words: Reflective professional inquiry; Reflective inquiry; Collaborative learning; Collaborative inquiry; Professional Learning Networks; Teacher professional development

Introduction

Educators collaboratively learning together is encouraged by school and school system leaders in the expectation that this will contribute to their ongoing teacher professional development (Butler and Schnellert, 2012; Pollard, 2005; Schleicher, 2012; Sebba *et al*, 2012). Furthermore, that such development should ultimately result in improved student outcomes (VanGrieken *et al*, 2017). For instance, in Professional Learning Networks, groups of educators (e.g., teachers, school leaders, educational researchers etc.) are facilitated to congregate and engage in collaborative learning focused on improving teaching and learning in participating schools (Poortman and Brown, 2018; Poortman, Brown and Schildkamp, submitted). Similarly, the idea of schools as professional learning communities

(PLCs), learning organisations (LOs), or even the concept of the school system as something that can be 'self-improving', envisions school leaders and teachers collaborating to jointly consider then share the leadership of instructional improvement (Stoll *et al.*, 2006; Greany, 2015; Kools and Stoll, 2016). In a recent systematic review, Doğan, and Adams (2018) conclude that facilitator/leader support, effective collaboration, a focus on instruction and students and reflective inquiry are all characteristics of the type of collaborative teacher learning associated with improvements in teachers' practice. Reflective inquiry is perhaps the most vital of these factors to 'get right' if teachers are to truly open themselves up to new ideas and ways of working as part of a collaborative endeavor focused on improvement. At the same time, reflective inquiry is currently poorly conceptualized in the literature (Poortman *et al.*, submitted).

That reflective inquiry can be considered poorly conceptualized would appear to be evident in the varying definitions and terms used to describe this form of teacher activity which may be associated with the complexity of reflection in itself (Jay and Johnson, 2002). Accordingly, notions of 'reflective dialogue', 'active learning strategies', 'reflective practice' and 'continuous teacher learning' (amongst others) have all been variously used to describe forms of reflective inquiry. At the same time, these terms all appear to represent (sometimes very) different processes (Doğan, and Adams, 2018; Lomos *et al.*, 2011; Meijlof, 2018; Pollard, 2005; Stoll *et al.*, 2006; Vangrieken, *et al.*, 2017; Vescio *et al.*, 2008). For the purpose of introducing this study, therefore, it is useful to provide some initial clarification of terms. As such, in considering notions of reflection we draw on Dewey (1933: 9), who describes it as 'active, persistent and careful consideration of any belief or supposed form of knowledge in the light of the grounds that support it and the further conclusions to which it tends'. Inquiry learning, meanwhile is thought of as representing a 'process of discovering new causal relations, with the learner formulating hypotheses and testing them by conducting experiments and/or making observations' (Pedaste *et al.*, 2015). Although we use these as initial working terms, one of the aims of this work, as we describe later, is thus to pull together extant thinking in this area in order to provide a working definition of RPI.

Moreover, understanding how best to achieve RPI is also vital, because many models for school improvement are premised on practitioners possessing the capacity to engage collaboratively and reflectively with new knowledge and ideas. This can be seen, for example, in those recent forms of continuous organizational improvement that are grounded in notions such as 'improvement science' (Bryk *et al.* 2015; Lewis 2015), 'design-based school improvement' (Mintrop 2016; Penuel *et al.* 2011) and evidence-informed practice (Cain *et al.*, 2019; Gorard *et al.*, 2019). School improvement in this vein, typically consists of problem solving as an iterative, evidence-based cycle of inquiry (Brown and Greany, 2018). Various versions of these cycles exist, but most invokes a sort of rationality that may roughly follow what Bransford and Stein (1993) call the 'IDEAL sequence'; which consists of identifying, defining and diagnosing, exploring alternatives, acting, and looking back on a problem. Furthermore, while Doğan, and Adams (2018) could also link facilitator/leader support, effective collaboration and a focus on instruction and students to improvements in student learning, this was not possible with reflective inquiry. It can therefore be considered the 'weakest link' in a chain that connects models of collaboratively learning to improved student outcomes.

Difficulties in achieving RPI

In educational settings, achieving RPI appears to be challenging for a number of reasons, however; three of which are identified by Mintop and Zumpe (2019) as especially salient. These are: 1) the required pace of educational decision making; 2) the complexity of the problems encountered by educators; and 3) the competition between different logics of action that impose different ways of reasoning on school leaders and teachers alike. These factors are explored in more detail below:

- 1) **Fast paced decision making:** Mintrop and Zumpe (2019) suggest that half of the activities of school leaders last less than ten minutes and only eight per cent exceed an hour. As a result, they argue that school leaders tend to be action-oriented and tend to have a dislike for reflective activities, which invariably take longer. A similar conclusion is found in Cain *et al.*, 2019, who go on to note that some of the errors that can be attributable to this type of *fast* or 'System 1' thinking (Kahnemann, 2011) include: practitioners jumping too readily to conclusions; practitioners tending to notice evidence that supports their existing beliefs, whilst ignoring what might be contradictory; and practitioners tending to overestimate the extent of their students' existing knowledge. Such patterns have been long acknowledged: reviewing the literature on teachers' decision-making, for example, Shavelson (1983: 32) found that teachers were reluctant to change their thinking and acting, 'even if they are not proceeding as well as expected'. Likewise, Chinn and Brewer (1993) found that teachers and students fight hard to hold onto previously held beliefs and will reject or distort new ideas that contradict them.
- 2) **The complexity of the problems encountered:** often the scope of problems faced by educational practitioners means that a 'rational-comprehensive' method of problem solving can be unwieldy (Mintrop and Zumpe, 2019). This reveals a core dilemma in real-world problem solving: how to respond effectively to problems that are too complex to understand comprehensively and too significant to consign to fast-thinking, trial and error (*ibid*). As a result, problems are often solved by practitioners 'muddling through' or adopting 'simple recipes' rather than any comprehensive analysis of the problem and potential solutions (Cohen *et al.* 1972; Haberfellner and Fenzl, 2017). Muddling through reduces the complexity of problem solving, but by its very nature, it is ineffective since it can limit the range of solution choices to what seems possible within a narrow horizon of precedent and the status quo. It is often best tempered therefore by an optimal-normal mix of theory, experience and rationality (Dror, 1964).
- 3) **Competition between different logics of action:** When school leaders seek to solve problems in their organizations, their reasoning will be influenced by underlying logics of action (March and Olsen, 2006; Jesscher-Rößler, 2020). Such logics of action will shape what aspects of the situation come into view, determine what goals are privileged, delineate where one draws the boundaries of agency, and also provide a lens through which one views suitable solutions (Mintrop and Zumpe, 2019; Jesscher-Rößler, 2020). Research on educational organizations in the tradition of institutional theory (e.g. Honig, 2006) argues that when seeking to solve problems school leaders often privilege legitimacy: i.e. acting according to public expectations of what is appropriate, over effectiveness (Mintrop and Zumpe, 2019).

As a result of these three factors, approaches to develop the capacity of teachers and school leaders to engage in RPI - for example in Research Learning Networks, Data Teams or Designed Based thinking workshops (Brown, 2018; Brown *et al.*, 2017; Mintrop and Zumpe, 2019) - can often fail in their attempts to facilitate System 2 decision making (Kahnemann, 2011), which is described as heuristic processing in social-psychological dual-process models (e.g. Gregoire, 2003). In other words, rather than engage in conscious reflection and analysis regarding problems and solutions, those participating in such 'programmes' can often resort to System 1 actions (heuristic processing, Gregoire, 2003) such as: 1) an overreliance on intuition and/or heuristics, which typically contain inherent biases that are not addressed; 2) problem solvers judging the likelihood of an outcome not by careful analysis of a situation but rather to 'the ease with which the mental operation of retrieval, construction, or association can be carried out' (Tversky and Kahneman 1973: 208); 3) an overreliance on assumptions, which may not hold, as a means of determining the cause of problems; and 4) practitioners jumping too quickly to action in order to satisfy a pressure to look decisive; which results in problems being solved in a 'ready-fire-aim'

pattern (Fullan, 2007). According to Kahnemann (2011) all of these actions serve to diminish rather than enhance decision making; in particular, because such behaviour tends to result in fuzzy problem perception and forestalls any meaningful analysis of alternative solutions. Furthermore, it requires considerable mental effort to overcome such errors and often problem solvers are tempted to avoid any conscious effort to do so (Cain *et al.*, 2019; Gregoire, 2003).

Finding out more about how to facilitate RPI

It would appear, therefore, that there is insufficient understanding as to: i) how to enhance the capacity of practitioners and educationalists to engage in RPI – especially in the light of the issues outlined above; and ii) what impact any improved capacity for RPI might have on teaching, school and ultimately, student outcomes. As such, our aim was to close this knowledge gap by exploring what is currently known in terms of how to both conceptualise and facilitate RPI. To address these knowledge gaps, we undertook an exploratory literature review to discover how others were thinking about RPI and how they were developing approaches that were successfully fostering it. Our review was guided by and designed to answer the following four research questions:

- 1) How is RPI conceptualised and how does it impact on practice and practice-related outcomes?
- 2) What concepts or programmes exist in relation to facilitating RPI?
- 3) How can we measure changes in RPI: how do we assess whether approaches to foster RPI have made a difference in terms of facilitating reflective inquiry and subsequently, student outcomes?
- 4) What evidence exists in terms of the impact of approaches to/programmes for facilitating RPI?

At the same time, the notion of RPI is both relevant to and encompasses a range of disciplines. In addition to Education, these include Business (for example, in terms of *Learning Organisations* and in relation to concepts such as *Knowledge Creation*: Nonaka and Takeuchi, 1995; Kools and Stoll, 2016), Psychology (e.g. the work Kahnemann (2011) on *decision making*); Medicine, Nursing and Social Care (e.g. Ng *et al.*, 2015; Winkel *et al.*, 2017) and Art and Design (de Botton, 2002; Norman, 2013). In order to ensure our understanding was grounded in the best possible understanding of RPI, which in turn, we sought to include this wide range of perspectives formed part of our review.

Our approach

To take into account the methodological and theoretical differences that exist between the various disciplines considered, we wished to include, our approach drew on the ‘Meta-Narrative Approach’ to conducting literature reviews (Wong *et al.*, 2013). The Meta-Narrative Approach is one that seeks ‘to illuminate a heterogeneous topic area by highlighting the contrasting and complementary ways in which researchers [from different academic disciplines] have studied the same or a similar topic’ (Wong *et al.*, 2013: 2). The key stages in Meta-Narrative reviews are (Greenhalgh *et al.* 2005; Wong *et al.* 2013):

- i. A stepwise search strategy: an initial scoping of the literature ‘led by intuition, informal networking and ‘browsing’, with a goal of mapping the diversity of perspectives and approaches’; this is followed by citation tracking for seminar conceptual papers; and finally, searching for empirical papers by electronic searching of key databases, hand searching of key journals and ‘snowballing’ (references of references or electronic citation tracking).
- ii. An appraisal phase for judging the relevance of each study to the review questions, as well as the quality of the evidence that exists.
- iii. A synthesis phase that identifies and presents the most important dimensions associated with RPI. This should present key concepts, theories, interventions, key actors, enablers and barriers and identified impacts associated with types of RPI: for example, what RPI is and how it affects, and has been shown to affect, teaching practice and student outcomes.

- iv. A discussion and recommendations phase that draws out key insights for practice.

The review process

Stage i) of the review involved the identification of seminal texts by all authors, as well as a wider search strategy involving the search engines Google Scholar and Web of Science. These search engines were selected due to their extensive use by researchers and their perceived complementarity (Martín-Martín *et al.*, 2019): Google Scholar is a wide-ranging search engine which includes a broader range of sources than just academic journals; including, for example, dissertations, books, and working papers. Web of Science. meanwhile is a human-curated database that focusses on journal articles which allows for more nuanced exploration, such as searching for review articles. The terms used to conduct the search were derived from an initial engagement with key texts (e.g. Doğan, and Adams, 2018; Lomos *et al.*, 2011; Meijlof, 2018; Mintrop and Zumpe, 2019; Pollard, 2005; Stoll *et al.*, 2006; Vangrieken, *et al.*, 2017; Vescio *et al.*, 2008). To narrow the search to relevant professional and informational areas, these terms were used in combination with four additional terms viewed as relevant to the study: '+ educat*', + teach*, + nurs*, + evaluati*. To provide an indication of what emerged from the search, the results from Web of Science are presented in Table 1, below (note that multiple words were put in quotation marks, while '*' denotes wildcard characters):

Table 1: Search findings from Web of Science

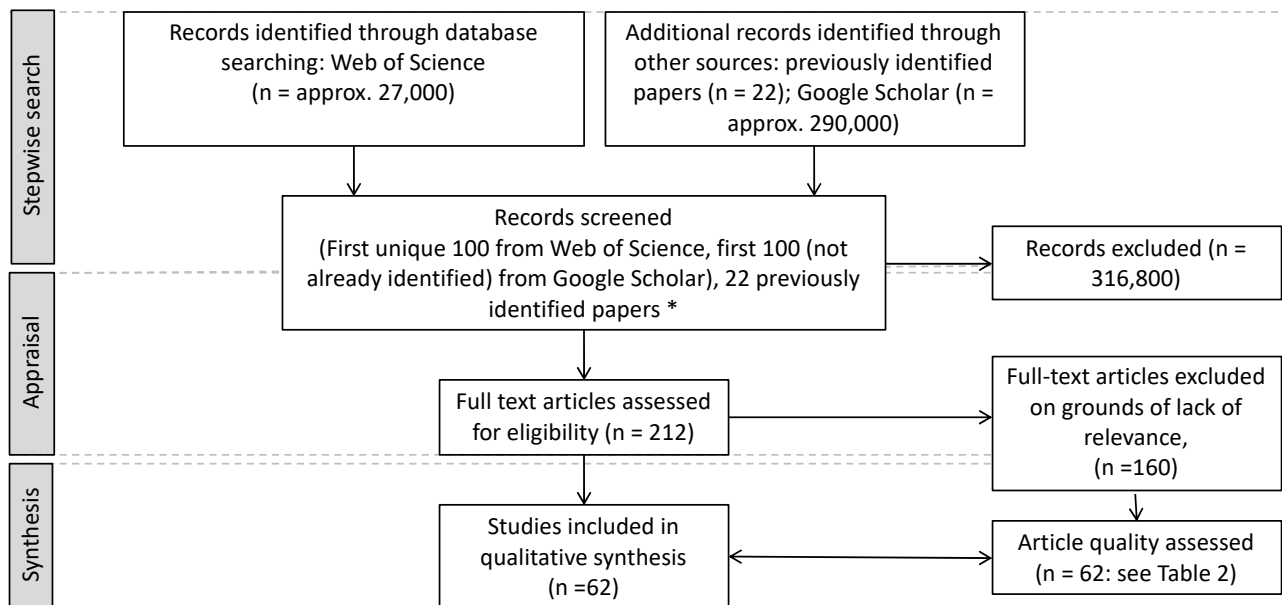
Web of Science	+ educat*	+ teach*	+ nurs*	+ result*	Total
Professional learning communit*	427	610	5	2	1,044
reflective inquiry	77	60	11	27	175
critical thinking	6,741	5,176	1,478	4,340	17,735
metacognition	1,516	1,213	53	2,987	5,769
framework for reflective thinking	1	1	0	1	3
learning conversation	15	18	2	11	46
learning network*	502	350	16	1284	2,152
networked learning communit*	17	7	0	4	28
theory in use	11	6	0	20	37
systematic reflection	48	36	7	34	125
teacher* reflective thinking	28	n/a	0	9	37
metacognition learning	4	2	0	4	10
reflective problem solving	4	3	0	3	10
Total	9,391	7,482	1,572	8,726	27,171

References from highly cited relevant papers, identified by authors one and two, were also accessed: for example, Rogers, 2002 which had been cited 1,938 times, Chinn and Brewer, 1993 which had been cited 2,088 times, Sparks-Langer and Colton 1991, cited 570 times. Stage i) highlighted that numerous researchers from a myriad of disciplines have examined ways to describe, encourage and measure RPI. As a result, the papers retrieved involved everything from existing literature reviews or critiques of existing literature, proposals for taxonomies of reflection and descriptions of situations where reflection has been assessed. Of particular note were the reviews of literature on general concepts relating to 'reflective thinking' published in the fields of nursing, medicine and education. These include: Akins and Murphy (1993), Baker (1996) and Platzer *et al* (1997), who reviewed literature in nursing; Koole *et al* (2011), Ng *et al* (2015) and Winkel *et al* (2017) who examined. reflection in medical education; and Fook

et al (2006), Mann *et al* in 2009 and Jayatilleke and Mackie (2013) who undertook reviews into reflection in health education. Reviews of literature in Education, included Hatton and Smith (1995), Rogers (2001), Dymont *et al* 2011, and Beachamp (2015). In addition, both Lee (2005) and Opfer and Pedder (2011) had reviewed literature on teacher professional learning; Stoll *et al.*, (2006) on Professional Learning Communities (PLCs) and Decuyper *et al.*, (2010) on team learning.

To gauge whether texts that emerged from the stepwise search warranted inclusion in the study, stage ii) involved the first and third authors each appraising 212 full text documents to assess topic relevance (based on alignment with the research questions) and rigour of method and reporting. There was broad agreement between authors one and three (with inter-rater reliability in the region of 90 percent), and where disagreements did occur in 19 instances, rationale for inclusion and exclusion were discussed. As a result of this process, there was a final agreement to exclude 160 studies from the review on the basis that they did not cover the topic criteria sufficiently; with 62 papers flagged for inclusion. A flow diagram providing further detail on the search strategy employed by the review is depicted in Figure 1:

Figure 1: A flow diagram detailing search strategy employed by the review



As noted above, stage ii) of the review also involved assessing the quality of the evidence our search strategy had uncovered. To do so, we drew on the research quality assessment framework developed by Gorard *et al.*, (2019). Gorard *et al.*, (2019) argue that the quality of research studies, and therefore the trustworthiness of their findings, should be judged based on the underlying research design of the project in question, the scale (i.e. size) of the sample, any missing data, the quality and relevance of the measurements taken, as well as fidelity and validity. The first step in using the framework is to identify the information on each of these quality factors from each paper in question. If the report does not include key information, or is written in such a way that the reader cannot ascertain this information, then Gorard *et al.* (2019) suggest that the research must be rated as having low security and doubt must be cast on its findings. Gorard *et al.*'s (2019) resulting five levels of quality are summarised in Table 2, below, with studies awarded a rating representing its lowest row description for any of the first four columns.

Table 2: A framework for assessing research quality (Gorard *et al.*, 2019)

Design	Scale	Completeness of data	Data quality	Rating
Strong design for research question	Large number of cases per comparison group	Minimal missing data, no evidence of impact on findings	Standardised, independent, pre-specified, accurate	4 <input type="checkbox"/>
Good design for research question	Medium number of cases per comparison group	Some missing data, possible impact on findings	Standardised, independent, not pre-specified, some errors	3 <input type="checkbox"/>
Weak design for research question	Small number of cases per comparison group	Moderate missing data, likely impact on findings	Not standardised, independent, or pre-specified, some errors	2 <input type="checkbox"/>
Very weak design for research question	Very small number of cases per comparison group	High level of missing data, clear impact on findings	Weak measures, high level of error, too many outcomes	1 <input type="checkbox"/>
No consideration of design	A trivial scale of study, or number is unclear	Huge amount of missing data, or not reported	Very weak measures, or accuracy not addressed	0 <input type="checkbox"/>

Applying the scales set out in Table 2 to the papers retrieved as part of the review revealed that not one achieved any score above ‘1 padlock’ level. The reason for this is that studies exploring the notion of RPI tended to be theoretical in nature, or were opaque, or of poor design in terms of their evaluation of the outcomes of RPI-related interventions; and this was the case irrespective of the research approach employed by such studies. The result, therefore, is that we did not come across one single RPI-related intervention that could be immediately adopted by PLN participants with any confidence that teaching practice or student outcomes would be positively impacted as a result. We return to the implications of this quality issue later, but note at this stage that we decided to include all papers deemed relevant to the review despite their low quality ratings. As a result, our findings and how we address the research questions are necessarily more theoretical than empirically based. Stage iii) of the review involved a process of synthesis to address the study’s four research questions and we present the result of this synthesis below, before we conclude with Stage iv): discussion and recommendations.

Stage iii) Synthesis of findings

The synthesised findings from our review are set out below by research question.

Research question 1: How is RPI conceptualised and how does it impact on practice and practice-related outcomes?

First and foremost, it should be noted that the review confirmed our earlier suggestion that there is no generally accepted definition of RPI. Moreover, that many terms have been used to describe what can broadly be considered the same concept, although there are differences in underlying theory and

emphasis. Overall, however, it can be suggested that the notion of RPI typically refers to the collaborative and co-regulative conversations practitioners have about serious educational issues or problems; with the collective generation and testing of ideas linked to the mutual scaffolding or support by colleagues as they jointly examine their practice (e.g. see Brown, 2020a; Butler and Schnellert, 2012; Kools and Stoll, 2016; Poortman and Brown, 2018; Stoll *et al.*, 2006). As such, achieving RPI would seem to involve practitioners actively and collectively questioning ineffective teaching routines, while finding proactive means to acknowledge and respond to them (e.g. Hubers and Poortman, 2018; Kools and Stoll, 2016; Thompson and Pascal, 2012). Effective RPI has been associated in the literature with notions such as 'reflective dialogue' - conversations about serious practice-related issues or problems involving the application of new knowledge in a sustained manner; the 'deprivatisation of practice' - frequent examining of practice, through mutual observation and case analysis, joint planning and development (e.g. of the curriculum); practitioners seeking new knowledge; practitioners' tacit knowledge constantly converted into shared knowledge through interaction; and practitioners applying new ideas and information to problem solving and solutions addressing the needs of, for example, students (Sebba *et al.*, 2012; Stoll *et al.*, 2006). As a result, RPI is considered more effective when it involves a high level of 'depth of inquiry'; with depth of inquiry defined as the degree to which practitioners employ higher level thinking skills when engaged in actions such as analysis, synthesis, goal setting, and reflection (Opfer and Pedder, 2011; Schildkamp and Datnow, 2020; Schildkamp *et al.*, 2016).

In education, when teachers display high depth of inquiry, they tend to be engaged in meaning making and the development new knowledge based on evidence and focused on the planning of actions to improve teaching and learning (Rogers, 2002). Practitioners that engage in lower levels of depth of inquiry are typically focused more on sharing information, storytelling and describing existing actions (Schildkamp and Datnow, 2020; Vanlommel *et al.*, 2017). In other words, with higher 'depth of inquiry', practitioners seek to further their understanding and turn this into action rather than simply transmit 'what is known'. Or as Warren-Little (1990) puts it, RPI should be directed towards the development of well-informed choices, rather than the mutual reinforcement of poorly informed habits. Furthermore, when practitioners display high levels of 'depth of inquiry', evidence is used conceptually and/or instrumentally as opposed to strategically or symbolically. Evidence is employed to change teachers' and school leaders' thinking (something referred to as conceptual evidence use) and used to make actual changes in classrooms and school (instrumental evidence use) (Farley-Ripple *et al.*, 2018; Schildkamp and Datnow, 2020). This division between conceptual and instrumental is also reflected by Sparks-Langer (e.g. Sparks-Langer *et al.*, 1990; Sparks-Langer and Colton 1991) who examine the use of reflection in the context of teacher training in the USA. Sparks-Langer argues that RPI can lead to both conceptual and instrumental impacts on teaching practice in three ways: technically (thinking about how to do something – the means); practically (which includes the means but also the ends – where do I want to be) and critically (which includes the technical and practical but also introduces moral and ethical dimensions). The notion of high 'depth of inquiry' is also associated with lower levels of external attribution and increased internal attribution. In particular low student achievement results are not automatically attributed to the external factors, such as the students' level of achievement when they enter the school, their social economic status or family backgrounds: reasons which do not typically require teachers to need to seek to change their practice (Datnow and Park, 2018). Rather, internal attribution leads practitioners examine their own functioning, challenging (often long held) assumptions and beliefs about student learning and looking to the quality of teaching for improvements in outcomes (Schildkamp and Datnow, 2020; Schildkamp *et al.*, 2016).

Research question 2: What concepts or programmes exist in relation to facilitating RPI?

A number of concepts and programmes emerged from the literature review in terms of how RPI can be facilitated. Such ideas include: i) learning conversations, ii) fostering cognitive dissonance, and iii) considering the role of emotion. These are set out in full below:

Learning conversations: RPI is thought to occur when teachers and other practitioners engage with a range of perspectives through open debate and discussion with others (Decuyper, *et al.*, 2010; Kools and Stoll, 2016; Lave and Wenger, 1991; Meijlof 2018; Nonaka and Takeuchi, 1995; Rogers, 2002; Schaap and de Bruijn, 2018; Stoll, 2010; Stoll *et al.*, 2006). Learning conversations foster teachers' collective questioning with regards to issues such as the school's fundamental purpose or regarding the implications of the pattern of practices that have accumulated over time (Horn and Warren-Little, 2010; Warren-Little, 1990). Learning conversations result in the 'creation' of new knowledge in the sense that multiple sets of knowledge are brought together to enable new understanding and development (Brown, 2017). To ensure knowledge creation is as effective as possible, high quality learning conversations are likely to feature the following characteristics (Decuyper, *et al.*, 2010; Meijlof 2018; Rogers, 2002; Schaap and de Bruijn, 2018; Stoll, 2010; Stoll *et al.*, 2006): 1) *a focus on evidence and/or ideas*: Learning conversations are focused on two important perspectives. First, an understanding of existing and effective practice within the school. Second, ideas about innovation and transformation where, for example, the conversation explores creative ways to engage learners and extend learning (Both foci require all those participating in the learning conversation to be committed to the problem area in question, which in itself will need to be linked to the problem area being focused on. Participants must also be prepared to offer and accept challenge and value the personal and intellectual growth of self and others; 2) *experience and external knowledge/theory*: access to outside expertise deepens learning conversations. Whether delivered personally, through writing, or via other media, independent ideas are injected to stimulate reflection, challenge the status quo and extend thinking. Such ideas can help promote greater depth in conversations; 3) *protocols and tools*: Learning conversations can often be framed more clearly when supported by frameworks and guidelines that help participants structure their dialogue and interrogate evidence or ideas. Teachers also need opportunities to look at and discuss 'artefacts' of their practice, not just test results; and 4) *facilitation*: facilitation is needed to ensure that the learning conversation approach is systematic, rigorous and disciplined. In other words, that that it elicits and supports intellectual exchange, as well as maintain open dialogue and, sometimes, injecting new energy into the conversation. Skilful facilitation can often lead to a productive balance of comfort and challenge. Time to enable learning conversations is key and also relates to both school leaders (who can afford teachers the time to engage) and the role of the facilitator in supporting learning conversations to happen (Brown 2017; Brown, 2020a; Huggins *et al.*, 2011).

Fostering cognitive dissonance: Chinn and Brewer (1993) describe the ways individuals change their beliefs (theories) in the presence of new information. Situating their study in the context of science teaching, they integrate knowledge from both psychology and science to present strategies to help students move their knowledge forward. In particular, their approach involves the introduction of anomalous data to challenge currently held theories. Responding to anomalous data involves coordinating existing beliefs and data; the new data needs to be examined to see whether it is believable, whether it can be explained within existing theories and if not whether or how the theory needs to be changed. Chinn and Brewer (1993) note that there is a tendency for people to shield their held beliefs and reluctantly give them up and suggest that seven responses tend to occur when individuals are faced with the introduction of anomalous data: it can be ignored, rejected, excluded from the theory domain, held in abeyance (to perhaps be dealt with at a later time), reinterpreted to fit the theory, the theory can be slightly changed to accommodate the data or the data can be accepted

and the theory changed. This last one, which involves the creation of cognitive dissonance, is the only one to change beliefs.

Models and simulation tools ('ready reckoners') can also be used to help create the cognitive dissonance amongst education practitioners that supports System 2 decision making and overcomes the difficulties associated with complexity, competing goals and the need to quickly engage in action (Tymms and Brown, submitted; Kahnemann, 2011). In particular, this is because such models help illustrate to those using them the difference between their theory in use and their espoused theory. First suggested by Argyris and Schön (1978), 'espoused theory' represents what people believe is their governing theory, 'theory in use', meanwhile, is how they actually behave (i.e. the theory that emerges from an individual's actions). More specifically, 'theory in use' describes the governing variables or values that the person holds and the action strategies that a person uses to keep their governing variables in a reasonable range. Argyris and Schön (1978) argue that when confronted by a situation where differences between 'espoused theory' and 'theory in use' are exposed, two strategies can be employed. In single-loop learning, an action strategy will be found that allows the underlying governing variable to remain uncontested. In double-loop learning, the conflict between theory-in-use and espoused theory leads to an examination and change in the governing variables and thus the theory-in-use. Models, because of how they provide instant responses to certain inputs and strategies, thus make it more likely that double-loop learning occurs, since unexpected or undesired outcomes will prompt users to engage almost instantaneously in (System 2 type) conscious reflection and analysis regarding problems and solutions. There is, to an extent, a convergence with the creation of cognitive dissonance and de Botton's (2002) analysis of the Victorian artist John Ruskin. Ruskin argued that the only way we can 'possess' beauty is by understanding it and making ourselves conscious of the psychological and visual factors responsible for it. In part this should come through the encouragement of drawing which, requires us to 'notice' specific details. In turn this *noticing* makes conscious why we find things (such as landscapes) beautiful and so, as a result, we can find explanations for our tastes. Acts of noticing thus require us to disrupt our conceptual models or to re-see what we have often regarded for years without thinking. Engaging in such acts may therefore lead to better alignments between our espoused theory and our theory in use (Argyris and Schön, 1978). Better noticing may also result in us rebalancing the extent to which we externally or internally attribute the causes of lower student achievement (Schildkamp and Datnow, 2020).

Suggestions for how practitioners might notice better include, drawing on the world of computer science where programmers explaining their code to others in order to ensure their thinking and logic is sound, explaining out loud to others why we engage in acts or actions. But it can also include the keeping of reflective diaries and/or the regular writing of journal entries. Studies in this area take one of two perspectives. First, studies may explore the use of journaling from an assessment point of view with journal articles coded from a taxonomy (no single taxonomy is used extensively, see next section below). This coding can be at word, sentence or article level (see Bell *et al*, 2011; Boenink *et al*, 2004; Fakude, 2003; Hatton and Smith, 1995; Kember *et al*, 2000; Kember *et al*, 2008; Rees *et al*, 2005). Studies can also explore journaling from a learner point of view, where the journey from descriptive text to reflection is described for one or more students showing what deeper reflection has done for them (e.g. Brooks *et al.*, 2010; Cho and Huang, 2014; Clarke, 2004; Dervent, 2013; Fazio, 2009; Harrison and Lee, 2011; Sánchez-Martí *et al.*, 2018; Shoffner, 2008; Ostorga, 2006; Wade, 1996). A number of the studies consulted for this review (e.g. Fakude, 2003; Shoffner, 2008; Wade, 1996) look at improving the reflective skills of trainee teachers (or nurses) by introducing or expanding the number of times where reflective journaling takes place. For example, by focussing at the level of the situation/lesson/critical incident or other activity happening on a regular basis, with the regularity

allowing deeper reflection to take place (e.g. Hatton and Smith, 1995). A key critique here however is whether participants are getting better at reflecting or at writing journals (Brooks *et al.*, 2010).

The role of emotion: The field of Art and Design provides a useful insight into how emotion might be used to facilitate (or indeed hinder) RPI. Specifically, leading design academic, Donald Norman (2013: 47) argues that ‘the emotional system is a powerful information processing system...that determines whether a situation is safe or threatening, whether something that is happening is good or bad, desirable or not.’ In tense and threatening situations, the emotional system will trigger the release of hormones that bias the brain in preparation for action. In calm, non-threatening situations, the emotional system triggers the release of hormones that bias the brain towards exploration and creativity (Norman, 2013). A positive emotional state is therefore ideal for reflective thought while a brain in a negative emotional state provides focus: precisely what is needed to maintain attention on a task and finish it. Too much of either, however, results in tunnel vision, where people are unable to look beyond a narrow range of options (Norman, 2013).. This perspective links nicely with the educational perspectives provided by Schildkamp and Datnow (2020: 18), who argue that in data-teams (a specific PLN focused on the use of data to foster school improvement), how practitioners view the purpose of the data team is vital: with data use efforts focused on accountability being far less fruitful than those focused on continuous improvement, or an explicit focus on equity, which are ‘more likely to lead to school policies and practices that expand students’ opportunities to learn’. Schildkamp and Datnow (2020) also link such outcomes to emotion and suggest that when teachers experience negative experiences with data use, such as shaming and blaming or feel that their time is being wasted, they are far less likely to be engaged. Positive experiences, on the other hand, (for example, working with a productive team that is delving deeply into learning) are likely to encourage teachers to become more engaged and in turn, more reflective (display higher levels of depth of inquiry). This suggests, first, that RPI is more likely to occur when it the focus of the collaborative learning activity is linked to what matters to teachers: for example, if the goal of the learning relates to specific students, social or educational agendas or other ethical or value-rich motivators (Mintzes 2013; Pollard, 2005; Runhaar *et al.*, 2010). On the other hand, RPI is less likely to occur when teachers are expected to harness reflective professional inquiry in the service of accountability. Second, it also seems apparent that experiences will colour how fondly teachers remember their interactions and whether they dedicate further work in the service of RPI outside of the collaborative activity. In other words, whether teachers leave the work of collaborative learning ‘in the room’ or take it away with them, back to their classrooms. As Norman (2013: 10) observes ‘great designers produce pleasurable experiences’: this means the design of RPI workshops themselves and the exercises designed to foster RPI must be pleasurable rather than unpleasant. This doesn’t equate to unchallenging or uncritical, but it does mean that challenge and critique is achieved in such a way that RPI is subsequently regarded as an action (or a state of mind) to be repeated. Communities of practice (Lave and Wenger, 1998) can also be helpful here as a way of fostering norms about what matters to different groups of practitioners who collaborate together.

Research question 3: How can we measure changes in i.e. RPI: how do we assess whether approaches to foster RPI have made a difference?

The review revealed that approaches to measuring RPI require researchers to consider the full range of cognitive or behavioural changes are that researchers are trying to assess. Magolda and Porterfield’s (1988) model of Epistemological Reflection, for example, describes a hierarchy of ‘ways of knowing’ and their associated patterns of reasoning. Ways of knowing are classified as *Absolute*, *Transitional*, *Independent* and *Contextual*. Moving from *Absolute* to *Contextual* results in the learner reflecting more, since as they are more open to the idea of different ways of understanding knowledge. In both Mezirow (1991) and Merriam (2004), three types of reflection are proposed: content reflection (thinking about

specific experiences or issues); process reflection (thinking about how to deal with the experience/issue); and premise reflection (which involves examining assumptions and beliefs). The review revealed a number of taxonomies that might be used to classify levels of RPI which address these three types of reflection to a greater or lesser extent. For example, Sparks-Langer *et al* (1990: 27) provides a seven-level framework for assessing reflective discourse in relation to both problem identification and the development of potential solutions. These seven levels are as follows:

- 1) No descriptive language used
- 2) Simple, layperson descriptions used
- 3) 'Events' labelled with appropriate terms (which considering Merriam, 2004, would encompass experiences and issues)
- 4) Explanation of these events with tradition or personal preference given as the rationale
- 5) Explanation of these events with principles or theories given as the rationale
- 6) Explanation of these events with principles/theories and consideration of context factors
- 7) Explanation of these events with consideration of ethical, moral, political issues

Sparks-Langer *et al.*'s (1990) taxonomy does not appear to include the questioning of one's own long-held assumptions and beliefs, however, but could be modified to incorporate this. Finally, Kember *et al* (2000) and Kember *et al* (2008) have developed a schema, based on Mezirow's (1991) work into transformational learning, for evaluating the level of reflection. Their approaches include the use a questionnaire (2000) and evaluating the level of reflection in written work (2008). A similar four-point scale is used in both approaches and employs categories of *Habitual Action*, *Understanding*, *Reflection*, and *Critical reflection*. No indication was provided in either paper by Kember and colleagues, however, as to the validity of the survey or the evaluation of written work, nor how widely it has been employed.

Nonetheless, while taxonomies and scales exist, approaches to measuring how practitioners have engaged in RPI have been methodologically problematic since they have either used subjective measurement tools, such as self-report, or have involved the labour intensive approach of recording and analysing teacher written and oral discourse (meaning they have been small scale in nature) (Daly and Stoll, 2018). For example, Boenink (2004) created an instrument to assess medical students undertaking a medical ethics course as part of their training as well as assessing whether the new curriculum was effective. The instrument presents short case studies (vignettes) on which students are asked to provide between two and ten reflections. These were then were scored according to Boenink's scale. While considered to be valid by Boenink (2004) the process is both resource intensive and one that is problematic to scale-up widely. Lastly, it should be noted that a number of papers (e.g. Aronson 2011; Cavanagh and Prescott, 2010; Harrison and Lee, 2011; Runhaar, *et al.*, 2010) also describe a mismatch between reflective thought taught in courses with the reality of classroom practice, perhaps casting doubt on whether taxonomies such as those presented above actually measure meaningful change.

Research question 4: What evidence exists in terms of the impact of approaches to/programmes for facilitating RPI?

Although, as we note above, there are myriad and programmes in terms of how RPI can be facilitated, the review did not identify any strong evaluation evidence in terms of which approaches to fostering RPI have an impact on practice and/or related outcomes. Where evaluation studies exist they tend to: 1) correspond to very short time scales (e.g. Cho and Huang, 2014; Clarke 2004; Derwent 2015; Şendağ and Odabaşı, 2009), where there can be little chance of impact; 2) involve case studies which, while having longer timescales, represent small scale self-report studies that consider individuals' responses as to how they feel about their level of reflection (e.g. Bell *et al.*, 2010; Cajkler 2015; Cavanagh and Prescott,

2010; Clarke, 2004; Huggins *et al.*, 2011; Mintzes 2013); and 3) involve cross sectional questionnaires which cannot provide before and after assessments (Boenink, 2004; Kember *et al.*, 2000; Runhaar *et al.*, (2010). There also appears to be no evidence at all linking RPI to student (or other final stakeholder) outcomes. This evidence gap leads us to conclude that there appears to be no strong evidence base in terms of how best to facilitate RPI.

Discussion

This review was undertaken based on our contention that there is currently little understanding as to: i) how the capacity of practitioners and educationalists to engage in RPI can be enhanced; and ii) what the impact any improved capacity for RPI might have on teaching, school and student outcomes. As a result, we sought to fill a number of key knowledge gaps related to facilitating RPI. These included: 1) understanding how RPI is conceptualised; 2) the concepts/programmes designed to support RPI; 3) how to assess whether approaches to fostering RPI have made a difference to students and teachers; and 4) the impact of concepts/programmes for facilitating RPI. While our review findings have been somewhat successful in starting to close these gaps, we also believe that our review illustrates that there is still much to understand about how to successfully facilitate RPI. To begin with, the review confirms our initial suggestion that there is no generally accepted definition of RPI. Providing such a definition is vital, however, because without defining what RPI is, we cannot proceed to try and facilitate it, nor assess whether RPI impacts positively on teaching and learning. In considering knowledge gaps 2 and 4, our review suggests there are no existing interventions designed to foster RPI that have been rigorously evaluated, for example via Randomised Control Trials or through quasi-experimental methods. This means there are currently no approaches that we could confidently employ as part of Professional Learning Networks (or indeed other forms of professional inquiry) with the expectation that practice or student outcomes will improve as a result.

At the same time, in seeking to address knowledge gap 3, we have found that more effective approaches associated with measuring RPI need to be established. First because measuring how practitioners have engaged in RPI tend to be problematic since they either use subjective measurement tools, such as self-report, or involve the labour-intensive approach of recording and analysing teacher discourse (meaning they have been small scale in nature) (Daly and Stoll, 2018). There is also little detail on what impact any improved capacity for RPI might have on teaching practice and student outcomes; itself made difficult because of these methodological problems (Daly and Stoll, 2018). These issues thus need to be addressed in conjunction with the development of approaches to facilitating RPI, to ensure that these new RPI interventions can be evaluated meaningfully. As we have already illustrated above, rubrics for measuring levels of and changes to RPI clearly do exist. Since these rubrics appear to be robust, the most pernicious problem, therefore, is how they can be applied objectively and in at a large enough scale that changes in RPI can be considered significant. Furthermore, how changes in RPI for teachers can be linked to student outcomes.

Conclusions

We conclude our review with the following recommendations. First, since a definition is required, we argue that RPI can be justifiably thought of as: *a collaborative, dialogic process in which educators both consider and aim to address pressing educational issues or problems. Such a process involves the collective generation and testing of ideas linked to enhancing their own practice; with these ideas based on evidence in the form of literature and/or data and displaying internal attribution.* Having defined RPI in this way, we also argue that the review highlights a number of promising ideas on how this form of RPI might be facilitated. Specifically, the review indicates that interventions designed to enable RPI need to: i) foster interactive learning conversations that have a high depth of inquiry, that involve the

introduction of new ideas and that also involve practical collaboration in relation to the design and testing of new teaching practices; ii) foster cognitive dissonance as part of these conversations to ensure existing beliefs and assumptions are challenged (making it more likely that evidence, data and new perspectives or approaches are considered); and iii) attend to the role of emotion to ensure there is motivation amongst educators to want to be reflective (in other words to ensure that the subject of reflection actually matters to teachers). Nonetheless at this stage these ideas, having not been the subject to rigorous testing remain just that. To take them forward we therefore recommend they are incorporated as part of an intervention that is tested, first for promise of impact and then more rigorously to ascertain whether this pilot intervention is more effective than other approaches that might foster RPI. Such effectiveness testing should be undertaken using experimental or quasi-experimental approaches.

Finally, while we do not have a definitive conclusion for how effective and objective ways of measuring RPI might be achieved, we highlight that a number of nascent ideas for significant and objective assessment of RPI – which are, at the same time, also cost effective – set substantive store in adopting machine learning/Artificial Intelligence methodologies (e.g. Brown, 2020b). A machine learning/AI approach would involve training algorithms to analyse and identify aspects of reflective and non-reflective dialogue and could offer a number of benefits. In particular, once trained, an RPI recognition algorithm should take just minutes to accurately and objectively process hours of recorded conversation and produce a report to indicate whether teachers who are engaged in collaborative learning activity are also doing so reflectively. This approach would thus provide a way of quickly and efficiently analysing hours of data on practitioners engaging in RPI (which would take a regular research team much longer to achieve and also involve the substantive costs of transcription). As a consequence, an AI approach could potentially provide an objective, large scale, quantitative basis for assessing the success of any new approaches to facilitating RPI. As such, we believe, should be investigated further.

References

- Anderson, T. and Shattuck, J. (2012) Design-Based Research: A Decade of Progress in Education Research, *Educational Researcher*, 41, 1, pp. 16-25.
- Argyris, C. & Schön, D. A. (1978) *Organizational learning: A theory of action perspective* (Reading, MA, Addison-Wesley).
- Aronson, L. (2011) Twelve tips for teaching reflection at all levels of medical education, *Medical teacher*, 33, 3, pp. 200-205.
- Baker, C. R. (1996) Reflective learning: a teaching strategy for critical thinking, *Journal of Nursing Education*, 35, 1, pp. 19-22.
- Beauchamp, C. (2015) Reflection in teacher education: issues emerging from a review of current literature, *Reflective Practice*, 16, 1, pp. 123-141.
- Bell, A., J. Kelton, N. McDonagh, R. Mladenovic and K. Morrison (2011), A critical evaluation of the usefulness of a coding scheme to categorise levels of reflective thinking, *Assessment & Evaluation in Higher Education*, 36, 7, pp. 797-815.

Boenink, A., A. Oderwald, P. De Jonge, W. Van Tilburg and J. Smal (2004) Assessing student reflection in medical practice. The development of an observer-rated instrument: reliability, validity and initial experiences, *Medical Education*, 38, 4, pp. 368-377.

Bransford, J. and Stein, B. (1993) *The IDEAL Problem Solver: A Guide to Improving Thinking, Learning, and Creativity*, (New York, NY, Worth).

Brooks, E., C. R. Harris and P. H. Clayton (2010) Deepening Applied Learning: An Enhanced Case Study Approach Using Critical Reflection, *Journal of Applied Learning in Higher Education*, 2, pp. 55-76.

Brown, C. (2017) Research Learning Communities: How The RLC Approach Enables Teachers To Use Research To Improve Their Practice And The Benefits For Students That Occur As A Result, *Research for All*, 1, 2, pp. 387-405.

Brown, C. (2018) Research Learning Networks: A case study in using in using networks to increase knowledge mobilisation at scale, in: Brown, C. and Poortman, C. (Eds) *Networks for learning: effective collaboration for teacher, school and system improvement*, (London, Routledge) (pp. 238-55).

Brown, C. (2020a) *The Networked School Leader: How to improve teaching and student outcomes using learning networks* (London, Emerald).

Brown C. (2020b) How Artificial Intelligence can help bring about Efficiencies in Teacher Training, invited talk to the *Durham University, AI in Education symposium*, London, 10 March, 2020.

Brown, C. and Greany, T. (2018) The Evidence-Informed School System in England: Where Should School Leaders Be Focusing Their Efforts?, *Leadership and Policy in Schools*, 17, 1, pp. 115-137.

Brown, C., Schildkamp, K. and Hubers, M. (2017) Combining the best of two worlds: a conceptual proposal for evidence-informed school improvement, *Educational Research*, 59, 2, pp. 154-172.

Bryk, A., Gomez, L., Grunow, A. LeMahieu, P. (2015) *Learning to Improve: How America's Schools Can Get Better at Getting Better* (Cambridge, MA: Harvard Education Press).

Butler, D. L., and Schnellert, L. (2012) Collaborative inquiry in teacher professional development, *Teaching and Teacher Education*, 28, 8, pp. 1206-1220.

Cain, T., Brindley, S., Brown, C., Jones, G. and Riga, F. (2019) Bounded decision-making, teachers' reflection, and organisational learning: how research can inform teachers and teaching, *British Educational Research Journal*, <https://doi.org/10.1002/berj.3551>

Cajkler, W., P. Wood, J. Norton, D. Pedder and H. Xu (2015) Teacher perspectives about lesson study in secondary school departments: A collaborative vehicle for professional learning and practice development, *Research papers in education*, 30, 2, pp. 192-213.

Cavanagh, M. and A. Prescott (2010) The growth of reflective practice among three beginning secondary mathematics teachers, *Asia-Pacific Journal of Teacher Education*, 38, 2, pp. 147-159.

- Chinn, C. and Brewer, W. (1993) The role of anomalous data in knowledge acquisition: a theoretical framework and implications for science instruction, *Review of Educational Research*, 63, 1, pp. 1-49.
- Cho, Y. H. and Y. Huang (2014) Exploring the links between pre-service teachers' beliefs and video-based reflection in wikis, *Computers in Human Behavior*, 35, pp. 39-53.
- Clarke, M. (2004) Reflection: Journals and Reflective Questions: A Strategy for Professional Learning, *Australian Journal of Teacher Education*, 29, 2, n2.
- Cohen, M., March, J. and Olsen, J. (1972) A Garbage Can Model of Organizational Choice, *Administrative Science Quarterly*, 17, 1, pp. 1-25.
- Daly, A. and Stoll, L. (2018), Looking back and moving forward: where next for networks of learning, in: Brown, C. and Poortman, C. (Eds) *Networks for learning: effective collaboration for teacher, school and system improvement*, (London, Routledge) (pp. 205-214).
- Datnow, A., and Park, V. (2018) Opening or closing doors for students? Equity and data use in schools., *Journal of Educational Change*, 19, 2, 131-152.
- De Botton, A. (2002) *The Art of Travel*, (London, Penguin Books).
- Decuyper, S., F. Dochy and P. Van den Bossche (2010) Grasping the dynamic complexity of team learning: An integrative model for effective team learning in organisations, *Educational Research Review*, 5, 2, pp. 111-133.
- Dervent, F. (2015) The effect of reflective thinking on the teaching practices of preservice physical education teachers, *Issues in Educational Research*, 25, 3, pp. 260-275.
- Dewey, J. (1933) *How we think: a restatement of the relation of reflective thinking to the educative process* (Boston, MA, Houghton Mifflin).
- Doğan, S., & Adams, A. (2018). Effect of professional learning communities on teachers and students: reporting updated results and raising questions about research design, *School Effectiveness and School Improvement*, 29, 4, pp. 634-659.
- Dror, Y (1964) Muddling Through - 'Science' or Inertia?, *Public Administration Review*, 24, 3, pp. 153-57.
- Dyment, J. E. and T. S. O'Connell (2011) Assessing the quality of reflection in student journals: A review of the research, *Teaching in Higher Education* 16, 1, pp. 81-97.
- Fakude, L. (2003) Journaling: A quasi-experimental study of student nurses' reflective learning ability, *Curatoris* 26, 2, pp. 49-55.
- Farley-Ripple, E., May, H., Karpyn, A., Tilley, K., and McDonough, K. (2018) Rethinking connections between research and practice in education: A conceptual framework, *Educational Researcher*, 47, 4, pp. 235-245.
- Fazio, X. (2009) Teacher development using group discussion and reflection, *Reflective Practice*, 10, 4, pp. 529-541.

Fook, J., S. White and F. Gardner (2006) Critical reflection: a review of contemporary literature and understandings, *Critical reflection in health and social care*, pp. 3-20.

Fullan, M. (2007) *The New Meaning of Educational Change* (New York, NY, Teachers College Press)

Gorard, S. Griffin, N., See, B-H and Siddiqui, N. (2019) *How can we get educators to use research evidence. A review of the best ways to get evidence into use from many areas of public policy* (www.lulu.com).

Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O, and Peacock R. (2005) Storylines of research in diffusion of innovation: a meta-narrative approach to systematic review, *Social Science & Medicine*, 61, 2, pp. 417-430.

Greany, T. (2015) How Can Evidence Inform Teaching and Decision Making across 21,000 Autonomous Schools? Learning from the Journey in England, in Brown, C: *Leading the Use of Research & Evidence in Schools*, (London, IOE Press) (pp. 11–29).

Gregoire, M. (2003) Is It a Challenge or a Threat? A Dual-Process Model of Teachers' Cognition and Appraisal Processes During Conceptual Change, *Educational Psychology Review*, 15, pp. 147–179.

Haberfellner, C. and Fenzl, T. (2017). The utility value of research evidence for educational practice from the perspective of preservice student teachers in Austria - A qualitative exploratory study, *Journal for educational research online*, 9, 2, pp. 69–87.

Harrison, J. K. and R. Lee (2011). Exploring the use of critical incident analysis and the professional learning conversation in an initial teacher education programme, *Journal of Education for Teaching*, 37, 2, pp. 199-217.

Hatton, N. and D. Smith (1995) Reflection in teacher education: Towards definition and implementation. *Teaching and Teacher Education*, 11, 1, pp. 33-49.

Honig, M (2006) Street-Level Bureaucracy Revisited: Frontline District Central-Office Administrators as Boundary Spanners in Education Policy Implementation, *Educational Evaluation and Policy Analysis*, 28, 4, pp. 357–83.

Horn, I. S., and Little, J. W. (2010) Attending to problems of practice: Routines and resources for professional learning in teachers' workplace interactions, *American educational research journal*, 47, 1, pp. 181-217.

Hubers, M. and Poortman, C., (2018) Establishing Sustainable School Improvement through Professional Learning Networks, in C. Brown and C. Poortman, C. (Eds) (2018) *Networks for learning: effective collaboration for teacher, school and system improvement*, (London, Routledge) (pp. 194 to 204).

Huggins, K. S., J. J. Scheurich and J. R. Morgan (2011). Professional learning communities as a leadership strategy to drive math success in an urban high school serving diverse, low-income students: A case study, *Journal of Education for Students Placed at Risk*, 16, 2, pp. 67-88.

Jay, J. K., and Johnson, K. L. (2002) Capturing complexity: A typology of reflective practice for teacher education, *Teaching and Teacher Education*, 18, 1, pp. 73-85.

Jayatilleke, N. and A. Mackie (2013) Reflection as part of continuous professional development for public health professionals: a literature review, *Journal of public health*, 35, 2, pp. 308-312.

Jesscher-Rößler, L., (2020) *Responsive Schulleitung. Leitungshandeln zwischen Transferstreben und Legitimitätsanspruch* (Unpublished dissertation. University of Innsbruck)

Kahneman, D. (2011) *Thinking, fast and slow* (London, Allen Lane).

Kember, D., D. Y. Leung, A. Jones, A. Y. Loke, J. McKay, K. Sinclair, H. Tse, C. Webb, F. K. Yuet Wong and M. Wong (2000) Development of a questionnaire to measure the level of reflective thinking, *Assessment & evaluation in higher education*, 25, 4, pp. 381-395.

Kember, D., J. McKay, K. Sinclair and F. K. Y. Wong (2008). A four-category scheme for coding and assessing the level of reflection in written work, *Assessment & Evaluation in Higher Education*, 33, 4, pp. 369-379.

Koole, S., T. Dornan, L. Aper, A. Scherpbier, M. Valcke, J. Cohen-Schotanus and A. Derese (2011). Factors confounding the assessment of reflection: a critical review, *BMC medical education*, 11, 1, pp. 104.

Kools, M., & Stoll, L. (2016) *What makes a school a learning organisation: a guide for policy-makers, school leaders and teachers*, available at: <https://www.oecd.org/education/school/school-learning-organisation.pdf>, accessed on 12 September, 2019.

Lave, J., and Wenger, E. (1998) *Communities of practice: Learning, meaning, and identity* (Cambridge, Cambridge University Press).

Lee, H.-J. (2005). Understanding and assessing preservice teachers' reflective thinking, *Teaching and Teacher Education*, 21, 6, pp. 699-715.

Lewis, C. (2015) What Is Improvement Science? Do We Need It in Education?, *Educational Researcher*, 44, 1, pp. 54–61.

Little, J.W. (1990) The Persistence of Privacy: Autonomy and Initiative in Teachers' Professional Relations, *Teachers College Record*, 91, 4, pp. 509-535.

Lomos, C., Hofman, R. H., & Bosker, R. J. (2011), Professional communities and student achievement—a meta-analysis, *School effectiveness and school improvement*, 22, 2, pp. 121-148.

Lundgren, H. and R. F. Poell (2016). On critical reflection: A review of Mezirow's theory and its operationalization, *Human Resource Development Review*, 15, 1, pp. 3-28.

Magolda, M. B. B. and W. D. Porterfield (1988) Assessing Intellectual Development: The Link between Theory and Practice, (ACPA Media Publication No. 47).

Mann, K., J. Gordon and A. MacLeod (2009). Reflection and reflective practice in health professions education: a systematic review, *Advances in health sciences education*, 14, 4, pp. 595.

March, J. and Olsen, J. (2006) The Logic of Appropriateness, in: Goodin, R., Moran, M. and Rein, M. (Eds) *The Oxford Handbook of Political Science*, (Oxford, Oxford University Press), DOI: 10.1093/oxfordhb/9780199604456.013.0024

Martín-Martín, A., Orduna-Malea, E., Thelwall, M. and Delgado-López-Cózar, E. (2019) Google Scholar, Web of Science, and Scopus: Which is best for me?, available at: <https://blogs.lse.ac.uk/impactofsocialsciences/2019/12/03/google-scholar-web-of-science-and-scopus-which-is-best-for-me/>, accessed on 20 April 2020.

Meijlof, I. (2018) *Reflective dialogue in professional learning communities*, (Enschede, University of Twente).

Merriam, S. B. (2004) The role of cognitive development in Mezirow's transformational learning theory, *Adult education quarterly*, 55, 1, pp. 60-68.

Mezirow, J. (1991) *Transformative dimensions of adult learning*, (San Francisco, CA, Jossey Bass).

Mintrop, R. (2016) *Design-Based School Improvement: A Practical Guide for Education Leaders*, (Cambridge, MA, Harvard Education Press).

Mintrop R. and Zumpe, E. (2019) Solving Real-Life Problems of Practice and Education Leaders' School Improvement Mind-Set, *American Journal of Education*, 125, 3 (May 2019), pp. 295-344.

Mintzes, J. J., B. Marcum, C. Messerschmidt-Yates and A. Mark (2013). Enhancing self-efficacy in elementary science teaching with professional learning communities, *Journal of science teacher education*, 24, 7, pp. 1201-1218.

Nonaka, I. and H. Takeuchi (1995) *The knowledge-creating company: How Japanese companies create the dynamics of innovation*, (Oxford, Oxford University Press).

Ng, S. L., Kinsella, E. A., Friesen, F. and Hodges, B. (2015). Reclaiming a theoretical orientation to reflection in medical education research: a critical narrative review, *Medical education*, 49, 5, pp. 461-475.

Nonaka, I. and H. Takeuchi (1995). The knowledge-creating company: How Japanese companies create the dynamics of innovation, (Oxford, Oxford university Press).

Norman, D. (2013) *The design of everyday things* (Cambridge, MA, MIT Press).

Opfer, V. D. and D. Pedder (2011) Conceptualizing teacher professional learning, *Review of Educational Research* 81, 3, pp. 376-407.

Ostorga, A. N. (2006) Developing teachers who are reflective practitioners: A complex process, *Issues in Teacher Education*, 15, 2, pp. 5-20.

Pedaste, M., Mäeots, M., Siiman, L. de Jong, T., van Riesen, S., Kamp, E., Manoli, C., Zacharia, Z. and Tsourlidaki, E. (2015) *Phases of inquiry-based learning: Definitions and the inquiry cycle*, *Educational Research Review*, 14, pp. 47-61.

Penuel, W., Fishman, B. Cheng, B. and Sabelli, N. (2011) Organizing Research and Development at the Intersection of Learning, Implementation, and Design, *Educational Researcher*, 40, 7, pp. 331–37.

Pollard, A. (2005) *Reflective teaching (2nd edition)* (London, Continuum Publishing).

Poortman, C and Brown, C. (2018), The importance of Professional Learning Networks, in: Brown, C. and Poortman, C. (Eds) *Networks for learning: effective collaboration for teacher, school and system improvement*, (London, Routledge) (pp. 10-19).

Poortman, C., Brown., C. and Schildkamp, K. (under review) Professional Learning Network processes and the link to student outcomes: a conceptual model of impact and research opportunities, submitted to *Journal of Teacher Education*

Rees, C. E., M. Shepherd and S. Chamberlain (2005). The utility of reflective portfolios as a method of assessing first year medical students' personal and professional development, *Reflective Practice*, 6, 1, pp. 3-14.

Rogers, R. R. (2001). Reflection in higher education: A concept analysis, *Innovative higher education*, 26, 1, pp. 37-57.

Rodgers, C. (2002) Defining reflection: Another look at John Dewey and reflective thinking, *Teachers College Record*, 104, 4, pp. 842-866.

Runhaar, P., K. Sanders and H. Yang (2010) Stimulating teachers' reflection and feedback asking: An interplay of self-efficacy, learning goal orientation, and transformational leadership, *Teaching and Teacher Education*, 26, 5, pp. 1154-1161.

Sánchez-Martí, A., Puig, M. S., Ruiz-Bueno, A. and Regós, R. A. (2018). Implementation and assessment of an experiment in reflective thinking to enrich higher education students' learning through mediated narratives, *Thinking Skills and Creativity*, 29, pp. 12-22.

Schaap, H. and de Bruijn, E. (2018) Elements affecting the development of professional learning communities in schools, *Learning Environments Research*, 21, 1, pp. 109-134.

Schildkamp, K and Datnow, A. (2020) When Data Teams Struggle: Learning from Less Successful Data Use Efforts, *Leadership and Policy in Schools*, DOI: 10.1080/15700763.2020.1734630.

Schildkamp, K., Poortman, C. L., and Handelzalts, A. (2016) Data teams for school improvement, *School Effectiveness and School Improvement*, 27, 2, 228–254.

Schleicher, A. (2012). *Preparing teachers and developing school leaders for the 21st century: Lessons from around the world*, (Paris, OECD Publishing).

Sebba, J., Tregenza, J. and Kent, P. (2012) *Powerful Professional Learning: A school leader's guide to joint practice development*, (Nottingham: National College for School Leadership).

Şendağ, S. and H. F. Odabaşı (2009). Effects of an online problem based learning course on content knowledge acquisition and critical thinking skills, *Computers & Education*, 53, 1, pp. 132-141.

- Shavelson, R. (1983) Review of research on teachers' pedagogical judgment, plans, and decisions, *Elementary School Journal*, 83, 392–415.
- Shoffner, M. (2008). Informal reflection in pre-service teacher education, *Reflective Practice*, 9, 2, pp. 123-134.
- Sparks-Langer, G. M. and A. B. Colton (1991). Synthesis of research on teachers' reflective thinking. *Educational Leadership* 48(6): 37-44.
- Sparks-Langer, G. M., J. M. Simmons, M. Pasch, A. Colton and A. Starko (1990). Reflective pedagogical thinking: How can we promote it and measure it? *Journal of Teacher Education*, 41, 5, pp. 23-32.
- Stoll, L. (2010). Connecting learning communities: Capacity building for systemic change, In A. Hargreaves, A. Lieberman, M. Fullan & D. Hopkins (Eds) *Second International Handbook of Educational Change*, (Dordrecht, Springer) (pp. 469-484).
- Stoll, L., Bolam, R., McMahon, A., Wallace, M. and Thomas, S. (2006) Professional Learning Communities: A Review of the Literature, *Journal of Educational Change*, 7, 4, pp. 221–58.
- Thompson, N. and J. Pascal (2012) Developing critically reflective practice, *Reflective practice*, 13, 2, pp. 311-325.
- Tversky, A. and Kahneman, D. (1973) Availability: A Heuristic for Judging Frequency and Probability, *Cognitive Psychology*, 5, 2. pp. 207–32.
- Tymms, P. and Brown, C. (submitted) Modelling Educational Systems: An interactive approach to integrating theory and policy, submitted to the *Journal of Educational Administration*
- Vangrieken, K., Meredith, C., Packer, T. and Kyndt, E. (2017). Teacher communities as a context for professional development: A systematic review, *Teaching and Teacher Education*, 61, pp. 47-59.
- Vanlommel, K., Van Gasse, R., Vanhoof, J., and Van Petegem, P. (2017) Teachers' decision-making: Data based or intuition driven?. *International Journal of Educational Research*, 83, pp. 75-83.
- 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, pp. 80-91.
- Wade, R. and Yarbrough, D. (1996) Portfolios: A tool for reflective thinking in teacher education? *Teaching and Teacher Education*, 12, 1, pp. 63-79.
- Winkel, A. F., Yingling, S., Jones, A. and Nicholson, J. (2017) Reflection as a learning tool in graduate medical education: a systematic review, *Journal of Graduate Medical Education*, 9, 4, pp. 430-439.
- Wong, G., Greenhalgh, T., Westhorp, G., Buckingham, J. and Pawson, R. (2013) RAMESES publication standards: meta-narrative reviews, *BMC Medicine*, 11, 20, pp. 1-15.