Lesson Study effectiveness for teachers' professional learning : a best evidence synthesis

Reference:
Willems Iris, Van den Bossche Piet.- Lesson Study effectiveness for teachers' professional learning : a best evidence synthesis
International journal for lesson and learning studies - ISSN 2046-8253 - (2019), p. 1-19
Full text (Publisher's DOI): https://doi.org/10.1108/IJLLS-04-2019-0031

Lesson Study Effectiveness for Teachers' Professional Learning

A Best Evidence Synthesis

JOURNAL: International Journal for Lesson and Learning Studies
VOL/ISSUE NO: 00/00
ARTICLE NO: 629760
ARTICLE TITLE: Lesson Study effectiveness for teachers' professional learning: a best evidence synthesis
AUTHORS: Iris Willems, Faculty of Social Sciences, University of Antwerp, Antwerp, Belgium, and Piet Van den Bossche, Department of Training and Education Sciences, Faculty of Social Sciences, University of Antwerp, Antwerp, Belgium and School of Business and Economics, Maastricht University, Maastricht, The Netherlands

Abstract

Purpose – The purpose of this paper is to gather evidence for the effectiveness of Lesson Study (LS) for teachers' professional learning.
Design/methodology/approach – A systematic review of relevant papers published between January 2010 and April 2018 was performed. To accomplish the best evidence synthesis, relevant studies were selected based upon well-chosen keywords and inclusion criteria. Afterwards, the quality of the remaining studies was assessed by using a critical appraisal checklist combined with a strength and weakness analysis. Data sources included relevant articles identified through digital searches on Education Research Information Centre and Web of Science, as well through snowball sampling and personal contacts. A total of five studies meeting inclusion criteria were retained for this review.
Findings – These five studies describe LS as a powerful professional development approach because of its positive impact on teachers' professional learning. A significant improvement or change for knowledge, skills, behaviour and beliefs has been reported among teachers.
Originality/value – Only studies with a design that shows high effectiveness were selected. The sample of relevant studies is currently small. As a consequence more large-scale and long-term research which illustrates the short and long-term effects of the LS approach on teachers' and students' learning is desirable.
Keywords - Systematic review, Teacher professional development, Effectiveness, Lesson study, Teachers' professional learning
Paper type - Literature review

Introduction

Japan is the cradle of Lesson Study (LS). Since the start of the public education system in 1872, it has been the model for teachers' professional development (PD) (Takahashi, 2014). LS became widely known since the publication of 'The Teaching Gap' (Stigler and Hiebert, 1999). In this seminal work the authors indicate that the high results for mathematics and science in Japan could be explained by the LS approach. As a result, LS was introduced as a PD method in various countries and continents (Xu and Pedder, 2014). Elliott and Ling (2011) quote in the editorial of the International Journal for Lesson and Learning Studies*: LS has been characterised as "a system for building and sharing practitioner knowledge that involves teachers in learning from colleagues as they research, plan, teach, observe, and discuss a classroom lesson"* (Lewis, Perry and Friedkin, 2009, p.142).

Indeed, LS teachers go through four phases of the quality circle 'study-plan-conduct-reflect' (Lewis and Hurd, 2011) (Figure 1). The LS circle starts with the study of the teaching material. In this phase teachers take the time to consult, read and study courses, relevant scientific articles, available curricula and other sources (Takahashi and Yoshida, 2004). They formulate lesson goals and research questions (Murata, 2011). Next, the teachers design and plan or revise a *research lesson*. Together the teachers plan, develop or choose an educational method that visualises the students' learning (Murata, 2011; Takahashi and McDougal, 2016). One teacher teaches the lesson under the supervision of a professional expert (Takahashi and McDougal, 2016), while the other team members have a live observation of the students (Murata, 2011). During this third step, the observers collect data about how the students think, reason, answer and solve problems. As a final step in the process, the observers discuss and reflect on how students have responded to the teaching and learning process, and how to implement the course (Lewis and Tsuchida, 1997).

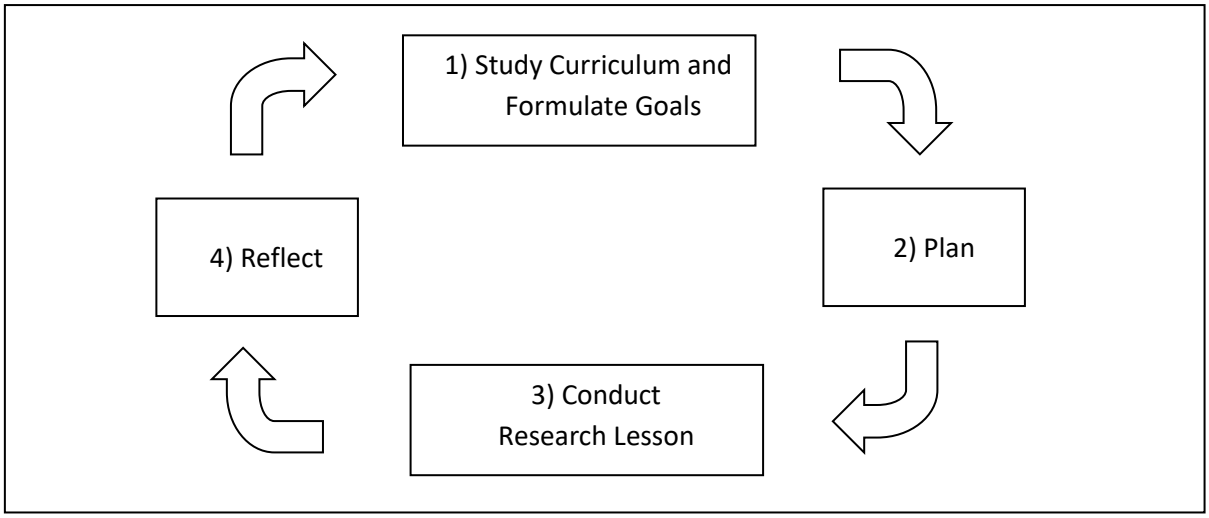

Figure 1. The Lesson Study circle (adapted from Lewis and Hurd, 2011, p.2).

However, LS is more than a series of steps to follow. The LS characteristics are the collaborative teacher teams, the centrally placed teaching practice, the lesson design they try out themselves, the learning processes of their own students within their own subject as a study object, and the long-term cyclical course (de Vries et al., 2017). These characteristics correspond to what recent reviews describe as effective and meaningful professionalization with a view to quality education (Darling-Hammond et al., 2009; van Driel et al., 2012).

LS has been researched and described in various countries over the last decades. Most of these studies are rather small-scale, qualitative, inductive and explorative (Xu and Pedder, 2014; de Vries et al., 2017). Cheung and Wong (2014) gathered for their review "Does Lesson Study work?", studies from 2000 till 2010 that demonstrate the effectiveness of LS. They find nine studies supporting LS benefits on instructional behaviour among teachers and on learning among pupils.

The existing literature shows that there is a need for well-controlled research that demonstrates the effectiveness of LS. Darling-Hammond et al. (2009) report: "Efforts to improve student achievement can succeed only by building the capacity of teachers to improve their instructional practice". LS is a

PD method with the aim of improving teachers' professional learning such as instructional practice (Lewis, 2002; Lewis et al., 2009). This review aims to gather evidence, for the period 2010 – 2018 (as a follow-up review of Cheung and Wong (2014, p.139), with regard to the research question: "Is Lesson Study effective for teachers' and teacher candidates' professional learning?"

Teachers' Professional Learning

LS has variants outside Japan that are often adapted and implemented as an innovative PD method in schools (Huang and Shimizu, 2016). Such educational innovations and changes often depend on the learning of teachers and on improving their instructions in order to achieve an increase in students' achievement (Matoba et al., 2007; Darling-Hammond et al., 2009).

To visualize this PD, Desimone (2009) has drawn up a conceptual framework (Figure 2). This framework allows researchers to map the effects of PD on both teachers and students, as recommended by Cheung and Wong (2014). It allows to investigate the following theories: 1) a theory about increased teacher knowledge and skills and changes of attitudes and beliefs, 2) a theory about changes in teacher instructions, for example a changed practice that affects students, 3) a theory that brings together both previous theories to gain insight into how PD functions. Desimone (2009) relies on existing literature to defend a number of relationships in the framework: (1) the link between PD and classroom practice (e.g., Fishman et al., 2003; Supovitz and Turner, 2000); (2) the link between PD and student performance (e.g., Angrist and Lavy, 2001; Cohen and Hill, 2000; Lee et al., 2008); (3) the link between teacher knowledge and practice, and students' performances (e.g., Hill, Ball and Schilling, 2008; Phelps and Schilling, 2004); (4) the link between teacher instructions and student performance (e.g., Hamilton et al., 2003; Supovitz, 2001; Von Secker, 2002). PD leads to changes and professional learning outcomes in terms of knowledge, skills, beliefs or behaviour. This improves the content of the teacher instructions and/or pedagogical approach (Desimone, 2009). Through processes of collective reflections during the LS cycle teachers develop new knowledge, understandings and beliefs about classroom practices and students' learning (Ricks, 2011).

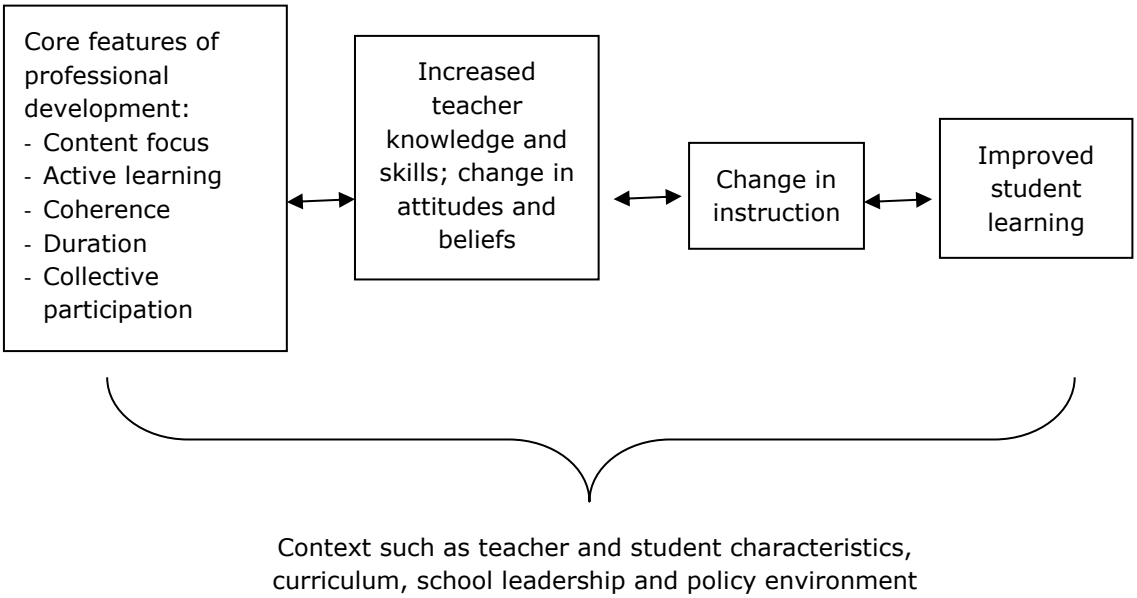

Figure 2. Proposed core conceptual framework for studying the effects of professional development on teachers and students (Desimone, 2009).

Because of the impact of teachers' knowledge, skills, attitudes and beliefs on student achievements (Desimone, 2009), this present review focuses on teachers, and by extension teacher candidates, as the target group, and teachers' professional learning during the LS intervention as the research subject. Teacher candidates are included, because teaching and education will become their future profession. The beliefs and practice of beginning teachers are often significantly influenced through pre-service teacher education programs (Kyles and Olafson, 2008). Research findings confirm also the effectiveness and impact of LS within initial teacher education (Lamb, 2015).

The current review, "Is Lesson Study effective for teachers' and teacher candidates' professional learning?", from January 2010 to April 2018, starts from the year where Cheung and Wong (2014) ended their review 'Does LS work?'. Furthermore it responds to their call for identifying evidence through randomized controlled trials (p.147).

Methodology

Best evidence synthesis

To accomplish the best evidence synthesis, relevant studies were selected based upon well-chosen keywords and inclusion criteria (see below). Afterwards, the quality of the remaining studies was assessed by using a critical appraisal checklist (CAC) combined with a strength and weakness analysis. Also information about differences in characteristics, outcome measures and their associated p-values, was used to evaluate the effectiveness of LS interventions on teachers' professional learning in terms of knowledge, skills, behaviour, and beliefs.

Keywords and inclusion criteria

A flowchart of the search process leading to the five studies that were analyzed in this systematic review is shown in Figure 3. The four main steps undertaken during this search process are identification, screening, eligibility, and inclusion. The number of articles involved in the data collection after each step is shown.

The digital databases Education Research Information Centre (ERIC) and Web of Science (WoS) generated at first 690 studies based on the search term "Lesson Study".

The second step involved Boolean searches in the digital databases resulting in 519 hits. The following keywords were individually combined with the main search term "Lesson Study": effect, quantitative, mixed method, posttest, outcome, impact, correlation, frequency, percentage, significant, statistic, survey, regression, control group, statistical analysis, t-test, ANOVA. This second step is aimed to focus on an effect measurement or the result thereof, a research design that allows an effect measurement or a measuring instrument that measures an effect. Removing doubles and not-peer-reviewed publications further reduced the number of articles from 519 to 212. Of these, 70 articles were considered suitable on the basis of three inclusion criteria: the main topic is LS, the publication date is from January 2010 to April 2018 and the target group is teachers or teacher candidates. Through snowball sampling four more studies were accepted at this stage.

The full-texts of these 74 articles were further assessed for eligibility in data reporting and suitability in design. According to What Works Clearinghouse (2008) only well-designed and well-implemented randomized controlled trials (RCT) can receive the highest rating of Meets Evidence Standards. The design had thus to be a RCT or a true/quasi-experimental design with control groups. A RCT is a trial with randomly assigned groups in order to determine the effectiveness of an intervention given to one of the two groups. RCTs provide the most reliable evidence for effects because the procedure greatly reduces the risk that the results are affected by disruptive factors. So RCTs are preferred in a systematic review to find out if an intervention is effective or not (What Works Clearinghouse, 2008).

Following this assessment eight studies remained. At this point, two additional articles were obtained through snowball sampling and one article through personal contacts.

Critical appraisal

These 11 studies were fully and thoroughly assessed using the critical appraisal checklist (CAC) (Appendix I). The purpose of this detailed assessment is to make a decision about the sound statistical approach in the published research and to retain articles that can provide an answer to the research question of the current review. Therefore the CAC was based on two existing tools: *CAT: A Randomized Controlled Trial* of The Joanna Briggs Institute (Tufanaru et al., 2017) and *How do I critically appraise the literature?* (Aveyard, 2014, p.114-121).

The description of the research context and participants, the LS intervention, the research design and instruments, the data collection procedure and data analysis were critically reviewed and recorded into the resulting dataset. During the data collection process, the details from the CAC were also identified as a weakness or as a strength of the research.

Ultimately, five articles were retained for this review: three articles obtained through the digital databases and two through snowball sampling.

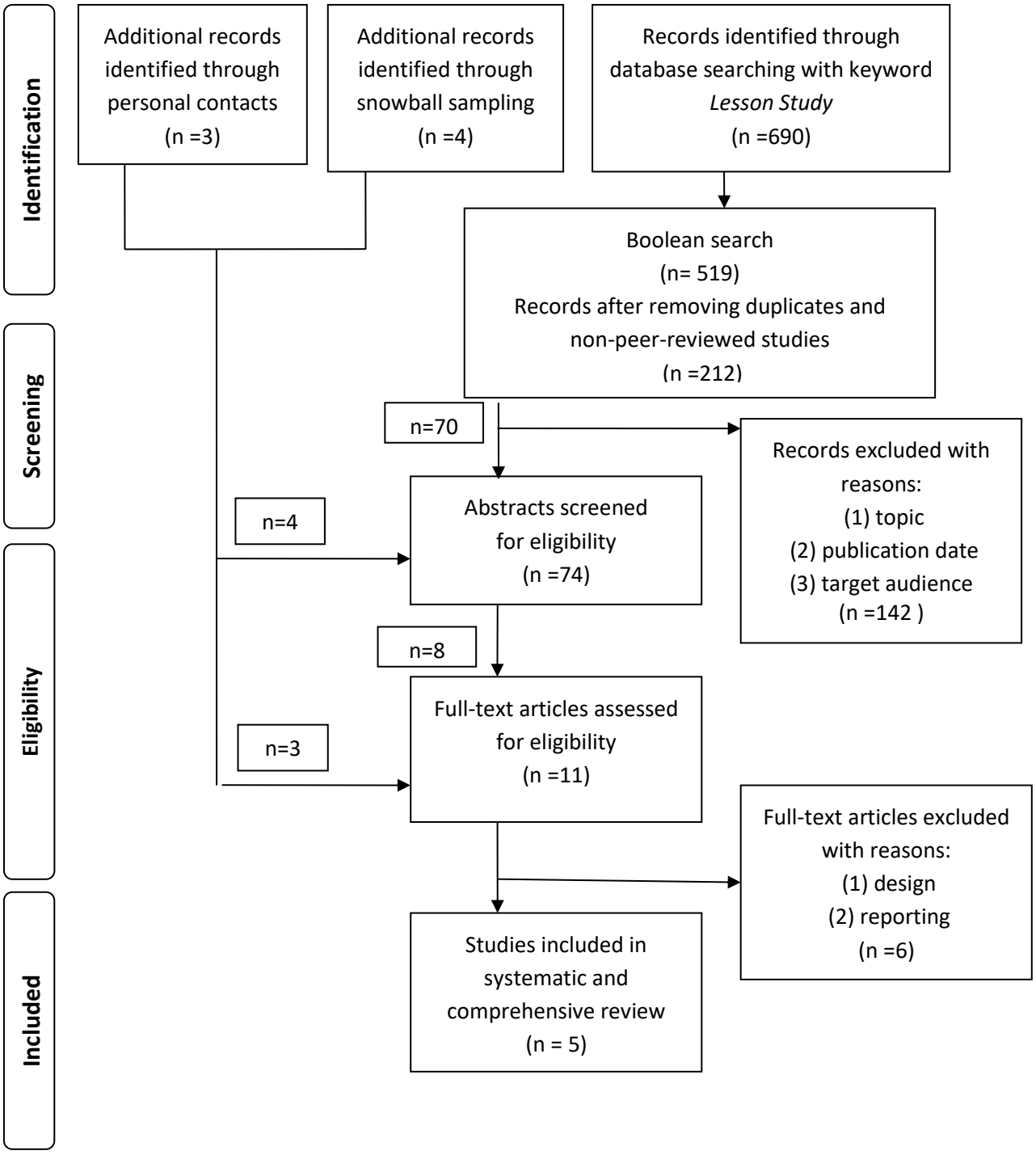

Figure 3. Flowchart of Screening Process with Results (adapted to Moher et al., 2009).

At this stage it appeared that two publications by Lewis and Perry (2015; 2017) used the same data set, conditions and research method. These two studies will therefore be reported together.

Data coding

The data were coded according to their major characteristics as represented in Table 1. The characteristics are defined as: type of education, geographical location, subject matter, participants, number of interventions and research duration, theoretical framework used in combination with the LS approach.

Exclusion for evidence

One study was excluded from evidence synthesis under the following conditions: the study has a score of eight or less on a maximum score of twelve on the CAC in combination with no maximum score on the inclusive criteria checklist.

Findings

Characteristics

The number of participants ranged from 48 to 213 (Table 1). In the American studies experienced teachers were recruited in elementary school (Lewis and Perry, 2015; 2017) and in middle school (Mutch-Jones et al., 2012), or teacher candidates at university (Chizhik et al., 2017). The Dutch study (Schipper et al., 2018) selected experienced teachers in secondary schools. Science and mathematics are in most studies represented as subject matter, except for Schipper et al. (2018) who include also language and social sciences. The LS interventions cover a period of three or twelve months and count one to three LS circles.

To map teachers' professional learning, four studies combine LS with a conceptual framework (Lewis and Perry, 2015; 2017; Chizhik et al., 2017; Schipper et al., 2018). Lewis and Perry (2015; 2017) and Schipper et al. (2018) rely on Desimone's model (2009). For the research at university Chizhik et al. (2017) combine the LS cycle with Rogoff's apprenticeship model (2014). One study does not utilise an additional theoretical framework and uses only the LS circle (Mutch-Jones et al., 2012).

Strengths/weaknesses analysis

The data analysis reveals strengths and restrictions that may have influenced the search for statistical effects.

Table 1

Characteristics of the Studies.

Studie	Type of education / Geographic location	Subject Matter	Number of Participants / Schools	Number of Interventions / Duration	Theoretical Framework
Mutch-Jones et al. (2012)	Middle school / USA	Science	n=73 / 10 schools	two / one academic year	--
Lewis and Perry (2015; 2017)	Elementary school / USA	Mathematics	n=213 / 27 schools	One / 3 months	Desimone's Theoretical Framework (2009)
Chizhik et al. (2017)	University education / USA	Science	n=60 / 1 university	three / one academic year	Rogoff's Super Vision Model (2014)
Schipper et al. (2018)	Secundary school / Netherlands	Languages, science, social science	n=48 / 8 schools	At least two / one academic year	Desimone's Theoretical Framework (2009)

Strengths

The five research teams use a research design with intervention groups which go through one or more LS circles and control groups which receive a traditional PD approach. Lewis and Perry (2015; 2017) apply a RCT, Mutch-Jones et al. (2012) use a true-experimental design and Chizhik et al. (2017) and Schipper et al. (2018) utilise a quasi-experimental design.

To increase the credibility and validity of the results triangulation has been applied in all retained studies. The baseline and effect measurements in the studies were performed with a large range of measuring instruments. These include: national tests (Chizhik et al., 2017; Lewis and Perry, 2015), surveys (Mutch-Jones et al., 2012; Lewis and Perry, 2015; 2017), online questionnaires (Schipper et al., 2018), open-ended assessments with standardized rubric coding (Mutch-Jones et al., 2012), video and/or audio recordings (Lewis et al., 2017; Mutch-Jones et al., 2012), field and reflection notes (Chizhik et al., 2017; Lewis et al., 2015; 2017), interviews (Chizhik et al., 2017; Schipper et al., 2018), lesson plan analyzes (Lewis and Perry, 2017), assessment scales (Lewis and Perry, 2017), observation checklist (Schipper et al., 2018).

The studies did control for baseline differences in intervention and control groups, and checked the validity and reliability of the instruments. All five studies were longitudinal. They run at least two LS interventions. The studies by Lewis et al. (2015; 2017) are large scaled (27 schools) with a big sample (n=213).

The data analyses are tailored to the design: Lewis and Perry (2015; 2017) and Mutch-Jones et al. (2012) use a hierarchical linear model to gain insight into nested groups. Schipper et al. (2018) use an explanatory sequential mixed method design to gain insight into patterns and effect differences between the groups, while Chizhik et al. (2017) use a combination of MANOVA, multiple t-tests and ANOVA.

All studies describe their statistic results and analyses extensively.

Most studies undertake actions to avoid bias and therefore increase the reliability of their results. Mutch-Jones et al. (2012) check for multicollinearity and engage four researchers to assess observations with validated rubrics. Lewis and Perry (2015; 2017) and Schipper et al. (2018) investigate the internal coherence of their questionnaires and remove items after a factor analysis. In two studies, experts assisted in refining the used measuring instrument (Chizhik et al., 2017, Mutch-Jones et al., 2012).

In order to visualize the impact of LS on teachers' professional learning all studies, except Mutch-Jones et al., (2012), use a theoretical model (Table 1).

Weaknesses

A number of limitations for construct validity were identified. In the study by Mutch-Jones et al. (2012) an intervention bias occurs because the LS cycle was not fully completed by the participants. The research team also reports the lack of a professional expert in the LS team, an unstructured coaching and a problematic final test. Both, Chizhik et al. (2016) and Lewis et al. (2017), use a measuring instrument with marginal scale reliability and reported it as problematic. Further, the intervention group, in the study by Mutch-Jones et al. (2016), differs significantly from the control group for two of the four dependent variables at baseline.

Examples of potential problems with external validity were found in the study by Schipper et al. (2018). The research team was confronted with a result bias. In their study, large differences occur between schools in terms of time allocation, the use of case pupils during the observations, and group composition.

There is a potential problem in internal validity in all studies, because no research team is completely blind to the participants. Having observed no statistically significant effects, Chizhik et al. (2017) suggested the presence of two confounding variables, namely the unique one-to-one relationship between a student and his mentor and the various school cultures between the internships. The studies of Schipper et al. (2018) and Lewis and Perry (2015; 2017) rely on their LS network to recruit teachers, for whom participation is voluntary. There is thus no random allocation of the participants.

Effects on Teachers' Professional Learning

As shown in Table 2 an improvement, change or growth is determined among teachers in teaching behaviour (Schipper et al., 2018), beliefs (Lewis et al., 2015; Schipper et al., 2018), skills (Chizhik et al., 2017; Lewis et al., 2017; Mutch-Jones et al., 2012) and knowledge (Lewis et al., 2017; Mutch-Jones et al., 2012).

Table 2

Lesson Study Research: Significant Effects on Teachers' Professional Learning.

Studies	Research Questions	Teachers' Professional Learning capture in	Significant Effects on Teachers' Professional Learning
Schipper et al. (2018)	Does LS influences teachers' sense of self-efficacy?	Beliefs	An increase for: Efficacy in Instructional Strategies. A difference between groups for: Efficacy in Pupil Engagement.
	Does LS influences adaptive teaching behaviour?	Behaviour	An increase for: Clarity of Instructions, Activating Learning, Teaching Learning Strategies. A difference between groups for: Efficient Classroom Management, Clarity of Instructions.
Lewis et al. (2015)	What is the impact of LS with mathematical resources on teachers' beliefs and teacher learning community?	Beliefs	A positive impact on: Collegial Learning Effectiveness, Expectations for Student Achievement.
(2017)	How is the quality of professional learning at the three conditions?	Skills	A higher rate from both LS groups (whether or not using the resources kit) than the non-LS group for: Leadership Skills, Sense of Inquiry.
	Do LS teachers, supported by mathematical resources, increase their mathematical knowledge for teaching fractions?	Knowledge	A positive impact on: teachers' mathematical knowledge for teaching fractions.
Chizhik et al. (2017)	Do teacher candidates, in LS program, perform better than the control group, on the edTPA test?	Skills	A difference between groups on two of 15 subscales: Planning to Support Varied Student Learning Needs; Analysis of Student Learning.
Mutch-Jones et al. (2012)	Are there content learning gains in teachers' knowledge of science content and learning challenges of students with learning disabilities.	Knowledge	/
	Are there teachers' process gains about the ability: to identify student learning challenges; to adapt an instructional plan; to develop accommodations for students with learning disabilities?	Skills	An increase between baseline and mid-point assessment for the ability: to adapt an instructional planning; to generate more accommodations for students with learning disabilities.

Behaviour and beliefs

Schipper et al. (2018) want to investigate whether the LS approach influences teachers' beliefs of self-efficacy and their adaptive teaching behaviour. A pretest and posttest questionnaire and observation instrument were used to capture teachers' behaviour and beliefs.

The adaptive teacher behaviour appears to change significantly positive for *Clarity of Instructions* ($t_{(25)}$=-2.28, p<.05), *Teaching Learning Strategies* ($t_{(25)}$=-2.52, p<.05) and *Activating Learning* ($t_{(25)}$=-2.93, p<.01). At the same time an ANOVA with repeated measurements results in statistical significant differences between the groups in teacher behaviour for *Efficient Classroom Management* ($F_{(1,46)}$=7.71,

p<.05) and *Clarity of Instructions* ($F_{(1,46)}$=6.62, *p*<.05) in favour of the intervention group. The study also reported a growth in *Adaptive Teaching*, but no statistical significant difference between groups could be found (Schipper et al., 2018).
Schipper et al. (2018) measure also a positive significant impact on teachers' beliefs in terms of sense of self-efficacy on the subscale *Efficacy in Instructional Strategies* ($t_{(25)}$=-2.64, *p*<.05). In addition, they establish a significant difference for *Efficacy in Pupil Engagement* ($F_{(1, 46)}$ = 4.51, *p*<.05) in favour of the LS group.

The studies of Lewis and Perry (2015; 2017) included three conditions: in condition 1, teams conduct LS using the fractions LS resource kit; in condition 2, teams choose their PD method, topic and material, and in condition 3, teams engage in LS on a "locally chosen LS intervention" without the fraction resource kit.
Lewis and Perry (2015) included 33 items in their survey that capture beliefs, interests and dispositions. The six scales to measure teachers' beliefs and learning community are: "Expectations for Student Achievement", "Using and Promoting Student Thinking", "Interest in Mathematics and Inquiry Stance", "Research Relevance for Practice", "Collective Learning Effectiveness" and "Professional Learning Community".
The study reports a statistical significant change, after the LS intervention in condition 1 with the mathematical resource kit, on four of six belief measures between pretest and posttest. Three positive significant changes found for: *Collegial Learning Effectiveness* ($t_{(72)}$=2,751, *p*<.01), *Research Relevance for Practice* ($t_{(72)}$=1,945, *p*<.10), and *Using and Promoting Student Thinking* ($t_{(72)}$=3,499, *p*<.01) and one is negative significant: *Professional Community* ($t_{(72)}$=2,414, *p*<.05).

Skills and Knowledge

In terms of skills, three studies report a positive significant effect (Chizhik et al., 2017; Lewis et al., 2017; Mutch-Jones et al., 2012).
Lewis and Perry (2015; 2017), as described before, included three conditions. When comparing the three conditions in order to find out the impact of the prescribed focus on fraction and the supply of the fractions resource kit on the professional learning quality, Lewis and Perry (2017) found that the LS groups (conditions 1 and 3) rated their experience regarding the quality for their professional learning significantly higher for leadership and research skills than the non-LS group (condition 2) ($t_{(206)}$=2.24, *p*<.05).
Chizhik et al. (2017) use the edTPA, a performance-based assessment for teachers, to examine teacher candidates' achievements and the effect on teaching performance after LS interventions. The edTPA has 15 subscales divided under three sections: "Planning for Instruction and Assessment", "Instruction and Engaging Students in Learning" and "Assessment". The LS group has higher means than the control group on 13 of 15 subscales about skills in planning, instructions and assessment. Chizhik et al. (2017) found under the section "Planning for Instruction and Assessment" a significant effect for the subscale *Planning to Support Varied Student Learning Needs* ($F_{(2,60)}$=4.30, *p*<.05) which implies for the LS group better skills and experiences of pupil diversity. Under the section "Assessment", a significant effect on the subscale *Analysis of Student Learning* ($F_{(2,60)}$=6.24, *p*<.05) was found.

Mutch-Jones et al. (2012) created three versions of a teacher assessment for the base-line, mid-point and final assessment. The assessments aimed to capture teachers' knowledge and skills. This included skills to adapt an instructional planning to meet science learning goals for all students in an inclusive

science classroom, skills to generate accommodations for students with learning disabilities, knowledge of science concepts and processes, and knowledge of the learning challenges of students with learning disabilities.
A significant effect for instructional planning adaptation in favour of the LS group was found on the intercept. At the baseline the control group had a value of 2.06 (p<.00), for the LS group the level was larger by a value .57 (p<.05). Further, 66% of the variation that lead to improvements in teachers' instructional planning adaptation, is explained by the LS intervention. For the ability to generate more adjustments for pupils with a learning disability, a significant effect for the LS group was found on the intercept by a value .94 (p<.05); for the control group by value 4.54 (p<.00). Also 5% is attributed to the LS intervention. Mutch-Jones et al. (2012) reported a significant difference between groups in favour of the LS team for teachers' knowledge of science concepts and processes (F= 8.421; p<.01), and for teachers' knowledge of students' learning challenges in inclusive science classroom (F=6.049; p<.05).

To assess knowledge as part of teachers' professional learning, Lewis and Perry (2017) combined condition 2 and 3 as a control group. The measure was standardized and the result of the LS intervention with resource kit gave the standardized mean difference (effect size) *.19* between the LS intervention with resource kit and the control group. The results showed a positive statistically significant impact of the LS intervention with resource kit on educators' mathematical knowledge for teaching fractions.

Conclusions based upon review

Xu and Pedder (2014) carried out a comprehensive review to categorise all available LS publications since 1999. Most of the gathered research was small-scaled, qualitative, and inductive of nature. In Cheung and Wong's review (2014), addressing studies published between 2000 and 2010, four studies reported a positive significant impact on teachers' professional learning. But, due to the lack of well-controlled and high-quality studies the evidence was insufficient. Qualitative research continues to provide evidence for the potential of the LS PD approach on teachers' professional learning (Norwich and Ylonen, 2013) or teacher candidates' learning (Meng and Sam, 2013; Zhou et al., 2017) in terms of knowledge and skills (Aimah et al., 2017; Dudley, 2013; Nami et al., 2016; Warwick et al., 2016), behaviour and beliefs (Cajkler et al., 2014; Bruce et al., 2016; Hadfield and Jopling, 2016; Schipper et al., 2017; Yakar and Turgut, 2017).

Addressing the period 2010-2018, the current review is a best evidence synthesis that contributes to the literature by providing a systematic review of studies which all use well-controlled designs to demonstrate the effectiveness of LS for teachers' professional learning. While this review contains only five studies that investigate the effects of LS, it still provides valuable new insights into the effects of LS on the learning process of teachers.

First, it shows that to map teachers' professional learning, LS is usually combined with a conceptual framework. Second, it indicates that LS can be implemented in all levels of education and is effective both for teachers and teacher candidates (Table 1).

The five retained studies provide significant evidence for the positive effects of LS in terms of knowledge (Mutch-Jones et al., 2012), skills (Chizhik et al., 2017; Lewis et al., 2017), teaching behaviour (Schipper et al., 2018) and beliefs (Lewis et al., 2015; Schipper et al., 2018) (Table 2). But, as described by Fernandez and Yoshida (2004), LS is an empty shell that must be filled with the knowledge and skills of the collaboratively working team members. According to de Vries et al. (2017, p.28) knowledge and skills and the completion of the four LS steps are promoting or impeding conditional factors. Two studies reported problems on this matter. First, Mutch-Jones et al. (2012) found no evidence for their theory that "*sharing of expertise has positive effects on the content and processes of knowledge and understanding of learning disabilities challenges".* They explain this by referring to a lack of professional expertise in the LS group and the fact that the LS cycle that was not fully completed. Second, Lewis et al. (2017) constructed three groups: one LS group with a research-based mathematical kit, one LS group without kit, and one non-LS group. Lewis found a significant improvement in teachers' knowledge, but only when research-based resources were available and used by the LS team.

Future directions

LS is the catalyst of professional learning (Lewis et al., 2017; Mutch-Jones et al., 2012; Rock and Wilson, 2005) and this learning should be mapped out for both teachers and students by making use of a theoretical framework.

Desimone (2009) describes professional teacher learning within the domains of knowledge, skills, attitudes and beliefs.

Mishra and Koehler (2006) in their TPACK-model break down teacher knowledge into technological, pedagogical and content knowledge, when integrating this knowledge within a certain educational context. Schipper et al. (2018) use a Dutch version of the Teachers' Sense of Efficacy Scale (Tschannen-Moran and Hoy, 2001) to measure different levels of teacher self-efficacy, as operationalized constructs for convictions or beliefs.

It is thus possible to further detail the professional teacher learning domain when addressing teacher learning from a different perspective. That is why we advise future researchers to refine and operationalize the constructs for measuring teachers' professional learning.

Beside this operationalization, LS research must identify circumstances and conditions for collective and individual learning among teachers, teacher candidates and students. Indeed, the current review focuses on how LS affects teachers' professional learning, while the ultimate aim of LS is to improve student achievements through the effectiveness of professional teacher development.

Despite the many positive classroom experiences and the sufficient evidence of the effectiveness of LS provided worldwide in small qualitative research, there is still a lack of significant evidence.

While acknowledging the complexity and associated challenges of performing well-controlled and well-implemented long-term studies that map the impact of LS on their educational context, we believe that researchers and governments should invest in this kind of long term research.

Appendix I Critical Appraisal Checklist for Controlled Trials

Article

Author

Year

Record Number

	Yes	No	Unclear	NA
1. Is the study published in an academic journal and is it peer-reviewed?	☐	☐	☐	☐
2. Is the research question clear and is it clear why the study was conducted?	☐	☐	☐	☐
3. Is the sample big enough?	☐	☐	☐	☐
4. Has the appropriate sample been obtained?	☐	☐	☐	☐
5. Does the study conduct a RCT or a quasi-/true-experimental design with control groups?	☐	☐	☐	☐
6. Are treatment groups similar at the baseline?	☐	☐	☐	☐
7. Are participants analyzed in the groups to which they were randomized?	☐	☐	☐	☐
8. Is the data collection method appropriate for the study design?	☐	☐	☐	☐
9. Is appropriate statistical analysis used?	☐	☐	☐	☐
10. Are outcomes measured in the same way for treatment groups?	☐	☐	☐	☐
11. Are outcomes assessors blind to treatment assignment?	☐	☐	☐	☐
12. Are outcomes measured in a reliable way?	☐	☐	☐	☐

Overall appraisal: Include ☐ Exclude ☐ Seek further info ☐

Comments (Including reason for exclusion)

References

Aimah, S., Ifadah, M. and Bharati, D.A.L. (2017), "Building teacher's pedagogical competence and teaching improvement through lesson study", *Arab World English Journal,* Vol. 8 No. 1, pp. 66-78.

Angrist, J. and Lavy, V. (2001), "Does teacher training affect pupil learning? Evidence from matched comparisons in Jerusalem public schools", *Journal of Labor Economics,* Vol. 19 No. 2, pp. 343-369.

Aveyard, H. (2014), *Doing a literature review in health and social care, a practical guide.* Open University Press, Berkshire, England.

Bruce, C., Flynn, T. and Bennett, S. (2016), "A focus on exploratory tasks in lesson study: The Canadian 'Math for Young Children' project", *ZDM Mathematics Education,* Vol. 48 No. 4, pp. 541-554.

Cajkler, W., Wood, P., Norton, J. and Pedder, D. (2014), "Lesson study as a vehicle for collaborative teacher learning in a secondary school", *Professional Development in Education,* Vol. 40 No. 4, pp. 511-529.

Cheung, W. and Wong, W. (2014), "Does lesson study work?: A systematic review on the effects of lesson study and learning study on teachers and students", *International Journal for Lesson and Learning Studies,* Vol. 3 No. 2, pp. 137-149.

Chizhik, E. W., Chizhik, A. W., Close, C. and Gallego, M. (2017), "SMILE (Shared Mentoring in Instructional Learning Environments): Effectiveness of a lesson-study approach to student-teaching supervision on a teacher-education performance assessment", *Teacher Education Quarterly,* Vol. 44 No. 2, pp. 27-47.

Cohen, D. and Hill, H. (2000), "Instructional policy and classroom performance: The mathematics reform in California", *Teachers College Record,* Vol. 102 No. 2, pp. 294-343.

Darling-Hammond, L., Wei, R.C., Andree, A., Richardson, N., and Orphanos, S. (2009), *Profesional learning in the learning profession: A status on teacher development in the United States and abroad,* National Staff Development Council and The School Redesign Network, United States .

de Vries, S., Roorda, G. and van Veen, K. (2017), *Lesson study: Effectief en bruikbaar in het Nederlandse onderwijs?,* Expertisecentrum Vakdidactiek Noord van de Lerarenopleiding van de Rijksuniversiteit, Groningen, Netherlands.

Desimone, L. (2009), "Improving impact studies of teachers' professional development: Toward better conceptualizations and measures", *Educational Researcher,* Vol. 38 No. 3, pp. 181-199.

Dudley, P. (2013), "Teacher learning in lesson study: What interaction-level discourse analysis revealed about how teachers utilised imagination, tacit knowledge of teaching and fresh evidence of pupils learning, to develop practice knowledge and so enhance their pupils' learning", *Teaching and Teacher Education,* Vol. 34, pp. 107-121.

Elliott, J. and Ling, L.M. (2011), "Editorial", *International Journal for Lesson and Learning Studies,* Vol. 1 No. 1.

Fernandez, C. and Yoshida, M. (2004), *Lesson study: A Japanese approach to improving mathematics teaching and learning,* Routledge, New York, NY.

Fishman, J., Marx, R., Best, S. and Tal, R. (2003), "Linking teacher and student learning to improve professional development in systemic reform", *Teaching and Teacher Education,* Vol. 19 No. 6, pp. 643-658.

Hadfield, M. and Jopling, M. (2016), "Problematizing lesson study and its impacts: Studying a highly contextualised approach to professional learning", *Teaching and Teacher Education,* Vol. 60, pp. 203-214.

Hamilton, L., McCaffrey D., Stecher, B., Klein, S., Robyn, A. and Bugliari, D. (2003), "Studying large-scale reforms of instructional practice: An example from mathematics and science", *Educational Evaluation and Policy Analysis,* Vol. 25 No. 1, pp. 1-29.

Hill, H., Ball, D. and Schilling, S. (2008), "Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students", *Journal for Research in Mathematics Education,* Vol. 39 No. 4, pp. 372-400.

Huang, R. and Shimizu, Y. (2016), "Improving teaching, developing teachers and teacher educators, and linking theory and practice through lesson study in mathematics: An international perspective", *ZDM Mathematics Education,* Vol. 48 No. 4, pp. 393-409.

Kyles, C. and Olafson, L. (2008), "Uncovering preservice teachers' beliefs about diversity through reflective writing", *Urban Education,* Vol. 43 No. 5, pp. 500-518.

Lamb, P. (2015), "Peer-learning between pre-service teachers: Embracing lesson study", *International Journal for Lesson and Learning Studies,* Vol. 4 No. 4, pp. 343-361.

Lee, O., Deaktor, R., Enders, C. and Lambert, J. (2008), "Impact of a multiyear professional development intervention on science achievement of culturally and linguistically diverse elementary students", *Journal of Research in Science Teaching,* Vol. 45 No. 6, pp. 726-747.

Lewis, C. (2002), "Does lesson study have a future in the United States?", *Nagoya Journal of Education and Human Development,* Vol. 1 No. 1, pp. 1-23.

Lewis, C. and Hurd, J. (2011), *Lesson study step by step: How teacher learning communities improve instruction,* Heinemann, New Hampshire.

Lewis, C. and Perry, R. (2015), "A randomized trial of lesson study with mathematical resource kits: Analysis of impact on teachers' beliefs and learning community", in Middleton, J., Cai, J. and Hwang, S. (Ed.), *Large-Scale Studies in Mathematics Education. Research in Mathematics Education.* Springer International Publishing, Switzerland, pp. 133-158.

Lewis, C. and Perry, R. (2017), "Lesson study to scale up research-based knowledge: A randomized, controlled trial of fractions learning", *Journal for Research in Mathematics Education,* Vol. 48 No. 3, pp. 261-299.

Lewis, C., Perry, R. and Friedkin, S. (2009), "Lesson study as action research", in Noffke, S. and Somekh, B. (Ed.), *The SAGE Handbook of Educational Action Research,* SAGE Publications Ltd., London, pp. 142-155.

Lewis, C., Perry, R. and Hurd, J. (2009), "Improving mathematics instruction through lesson study: A theoretical model and North American case", *The Journal of Mathematics Teacher Education,* Vol. 12 No. 4, pp. 285-304.

Lewis, C. and Tsuchida, I. (1997), "Planned educational change in Japan: The shift to student-centered elementary science", *Journal of Educational Policy,* Vol. 12 No. 5, pp. 313-331.

Matoba, M., Shibata, Y. and Sakar Arani, M. (2007), School-university partnerships: A new recipe for creating professional knowledge in school", *Educational Research for Policy and Practice,* Vol. 6 No. 1, pp. 55-65.

Meng, C. and Sam, L. (2013), "Developing pre-service teachers' technological pedagogical content knowledge for teaching mathematics with the geometer's sketchpad through lesson study", *Journal of Education and Learning,* Vol. 2 No. 1, pp. 1-8.

Mishra, P. and Koehler, M. (2006), "Technological pedagogical content knowledge: A framework for teacher knowledge", *Teachers College Record,* Vol. 108 No. 6, pp. 1017-1054.

Moher, D., Liberati, A., Tetzlaff, J. and Altman, D. (2009), "Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement", *PloS Med,* Vol. 6 No. 7.

Murata, A. (2011), "Introduction: Conceptual overview of lesson study", in Hart, L., Alston, A. and Murata, A. (Ed.), *Lesson Study Research and Practice in Mathematics Education,* Springer, Dordrecht, pp. 1-12.

Mutch-Jones, K., Puttick, G. and Minner, D. (2012), "Lesson study for accessible science: Building expertise to improve practice in inclusive science classrooms", *Journal of Research in Science Teaching,* Vol. 49 No. 8, pp. 1012-1034.

Nami, F., Marandi, S. and Sotoudehnama, E. (2016), "CALL teacher professional growth through lesson study practice: An investigation into EFL teachers' perceptions", *Computer Assisted Language Learning,* Vol. 29 No. 4, pp. 658-682.

Norwich, B. and Ylonen, A. (2013), "Design based research to develop the teaching of pupils with moderate learning difficulties (MLD): Evaluating lesson study in terms of pupil, teacher and school outcomes", *Teaching and Teacher Education,* Vol. 34, pp. 162-173.

Phelps, G. and Schilling, S. (2004), "Developing measures of content knowledge for teaching reading", *CPRE Journal Articles,* 5.

Ricks, T. (2011), "Process reflection during Japanese lesson study experiences by prospective secondary mathematics teachers", *The Journal of Mathematics Education,* Vol. 14 No. 4, pp. 251-267.

Rock, T. and Wilson, C. (2005), "Improving teaching through lesson study", *Teacher Education Quarterly,* Vol. 32 No. 1, pp. 77-92.

Rogoff, B. (2014), "Learning by observerving and pitching in to family and community endeavors: An orientation", *Human Development,* Vol. 57 No. 2-3, pp. 69-81.

Schipper, T., Goei, S.L., de Vries, S. and van Veen, K. (2017), "Professional growth in adaptive teaching competence as a result of lesson study", *Teaching and Teacher Education,* Vol. 68, pp. 289-303.

Schipper, T., Goei, S.L., de Vries, S. and van Veen, K. (2018), "Developing teachers' self-efficacy and adaptive teaching behaviour through lesson study", *International Journal of Educational Research,* Vol. 88, pp. 109-120.

Stigler, J. and Hiebert, J. (1999), *The Teaching Gap: Best Ideas from the World's Teachers for Improving Education in the Classroom,* Free Press, New York, NY.

Supovitz, J. (2001), "Translating teaching practice into improved students performance", in Fuhrman, S. (Ed.), *From the Capitol to the Classroom: Standards-based Reform in the States. 100th Yearbook of the National Society for the Study of Education, Section 2,* University of Chicago Press, Chicago, pp. 81-98.

Supovitz, J. and Turner, H. (2000), "The effects of professional development on science teaching practices and classroom culture", *Journal of Research in Science Teaching,* Vol. 37 No. 9, pp. 963-980.

Takahashi, A. (2014), "Supporting the effective implementation of a new mathematics curriculum: A case study of school-based lesson study at Japanese public elementary school", in Li, Y and Lappan, G. (Ed.), *Mathematics Curriculum in School Education. Advances in Mathematics Education, Springer, Dordrecht,* pp. 417-441.

Takahashi, A. and McDougal, T. (2016), "Collaboration lesson research: Maximizing the impact of lesson study", *ZDM Mathematics Education,* Vol. 48 No. 4, pp. 513-526.

Takahashi, A. and Yoshida, M. (2004), "Ideas for establishing lesson study communities", *Teaching Children Mathematics,* Vol. 10 No. 9, pp. 436-443.

Tschannen-Moran, M. and Hoy, W. (2001), "Teacher efficacy: Capturing an elusive construct", *Teaching and Teacher Education,* Vol. 17 No. 7, pp. 783-805.

Tufanaru, C., Munn, Z., Aromataris, E. Campbell, J. and Hopp, L. (2017), "Chapter 3: Systematic reviews of effectiveness", in Aromataris, E. and Munn, Z. (Ed.), *Joanna Briggs Institute Reviewer's Manual,* The Joanna Briggs Institute, available from https://reviewersmanual.joannabriggs.org/.

van Driel, J., Meirinck, J., van Veen, K. and Zwart, R. (2012), "Current trends and missing links in studies on teacher professional development in science education: A review of design features and quality of research", *Studies in Science Education,* Vol. 48 No. 2, pp. 129-160.

von Secker, C. (2002), "Effects of inquiry-based teacher practices on science excellence and equity", *The Journal of Educational Research,* Vol. 95 No. 3, pp. 151-160.

Warwick, P., Vrikki, M., Vermunt, J., Mercer, N. and van Halem, N. (2016), "Connecting observations of student and teacher learning: An examination of dialogic processes in lesson study discussions in mathematics", *ZDM Mathematics Education,* Vol. 48 No. 4, pp. 555-569.

What Works Clearinghouse (2008), *Procedures and Standards Handbook (Version 2.1), a*vailable from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_v2_1_standards_handbook.pdf.

Xu, H. and Pedder, D. (2014), "Lesson study: An international review of the research", in Dudley, P. (Ed.), *Lesson study, professional learning for our time,* Routledge, London/New York, pp. 29-58.

Yakar, Z. and Turgut, D. (2017), "Effectiveness of lesson study aproach on preservice science teachers' beliefs", *International Education Studies,* Vol. 10 No. 6, pp. 36-43.

Zhou, G., Xu, J. and Martinovic, D. (2017), "Developing pre-service teachers' capacity in teaching science with technology through microteaching lesson study approach", *EURASIA Journal of Mathematics Science and Technology Education,* Vol. 13 No. 1, pp. 85-103.