Submitted to the Department of Educational Sciences of the University of Potsdam/

Eingereicht bei der Humanwissenschaftlichen Fakultät der Universität Potsdam

Critical Research Needs in Lesson Study: Then, Now, and Looking Forward

In partial fulfillment of the requirements for the degree of/

Kumulative Dissertation zur Erlangung des akademischen Grades

Doctor of Philosophy (Dr. phil.)/

Doktor der Philosophie (Dr. phil.)

by/von

Klara Kager, M.A.

Potsdam, 03.02.2023

Supervisors/ Betreuerinnen:
Prof. Dr. Miriam Vock
Dr. Anne Jurczok

Reviewers/ Gutachterinnen:
Prof. Dr. Miriam Vock
Prof. Dr. Anne Sliwka

Date of PhD Defense/
Datum der mündlichen Prüfung: 03.07.2023

Published online on the
Publication Server of the University of Potsdam:
https://doi.org/10.25932/publishup-60271
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-602711

Acknowledgments

First and foremost, I am grateful to my supervisors. Prof. Dr. Miriam Vock, thank you for your advice and guidance on every step of the way, thank you for asking tricky questions, and for always listening to my ideas. Dr. Anne Jurczok, thank you for your endless encouragement and for always taking time out of your busy schedule for me. I also thank Prof. Dr. Anne Sliwka for agreeing to review this dissertation.

I could not have written this dissertation without my amazing team. Thank you, Dr. Eva Kalinowski, for your critical eye and catching all the mistakes. Swantje and Julian, it is an honor to share an office with you—you made this experience so enjoyable and fun and I've learned a lot from both of you. A special thank you goes to Ina, because without Ina, the University of Potsdam (or at least I) would break down. I'd also like to recognize our Hiwis for their great support. Many thanks to Ben Fergusson, who makes me sound much smarter and articulate than I actually am and taught me valuable lessons about n- and m-dashes. I also want to thank Prof. Dr. Hartung-Beck for her valuable advice.

I am greatly indebted to Dr. John Mynott from the University of Aberdeen. Thank you for the insightful conversations, your feedback and encouragement, and for hosting me in Aberdeen. When you told me to "just write it" and to do it "till next Wednesday", that was exactly what I needed to hear. I am looking forward to collaborating with you in the future!

I also want to thank the teachers at our project schools for their collaboration and feedback. Thank you to the Lesson Study community and the great people I met at the WALS conference, who share my enthusiasm for research and education.

Lastly, I want to thank my family for their wonderful support from afar and for accepting that I now refer to a "Teesackal" as a "Teebeutel". Thank you to my friends, back home and here in Potsdam, for keeping my motivation high during this process. And, of course, I could not have done this without your support and encouragement, Effrosyni. Thank you for not letting me quit as soon as it got hard, thank you for always believing in me, thank you for persuading me that we needed a second computer screen at home to deal with all the excel files (we really did), and thank you for introducing me to Spotify (a key factor to the success of this dissertation). Thank you for the incredible fun we had working on our dissertations during lockdown despite only having one proper work desk in our flat. You are simply the best.

Summary

Lesson Study (LS) is a collaboration-based and iterative professional development (PD) approach that is rooted in the Japanese system of teacher education (Chokshi & Fernandez, 2004). A group of teachers identifies a research question relevant to their practice and co-plans a lesson that targets this question. While one teacher delivers the lesson, the other group members observe how students learn and take detailed notes. In a last step, teachers come together in a post-lesson discussion to reflect together on students' learning in response to the lesson and translate their observations into future pedagogical intentions (Lewis et al., 2019). LS therefore incorporates multiple features that have been identified in the research literature as integral to effective PD (e.g., Darling-Hammond et al., 2017; Lipowsky & Rzejak, 2015). Specifically, LS is a long-term process that consists of subsequent inquiry cycles, it is site-based and embedded in teachers' practice, it stimulates collaboration and reflection, focuses on student learning, and typically includes external experts who facilitate the process or provide additional insights (Lewis et al., 2019; Murata, 2011). Empirical findings have connected participation in LS to an increase in teachers' knowledge (e.g., Coenders & Verhof, 2019), their awareness of students' needs (e.g., Dudley, 2013), as well as their self-efficacy (e.g., Schipper et al., 2018). For all these reasons, LS is considered an attractive concept to support teachers' professionalization throughout their career. Since the turn of the 21st century, LS has rapidly gained international popularity and is currently practiced in over 40 countries around the world (Yoshida et al., 2021).

The spread of LS has gone hand in hand with the emergence of a research field that aims to investigate the efficacy of LS on teacher learning and explore the conditions and processes that make LS effective in diverse contexts. In 2006, shortly after LS was first transferred to the United States and gained popularity outside of Japan, Lewis and colleagues proposed three critical research needs to guide the growing research efforts into LS. These research needs were (1) the development of a descriptive knowledge base on LS, (2) the investigation of the processes through which teachers learn in LS, and (3) the use of design-based research cycles to study and improve LS. The review by Yoshida et al. (2021) indicates that the field has since accumulated an impressive knowledge base on LS. This knowledge base, however, primarily consists of small-scale, qualitative, and heavily contextualized research, which makes it challenging to synthesize and replicate findings (e.g., Hadfield & Jopling, 2016; Xu & Pedder, 2014). There is a consensus among scholars that the field has not yet generated definitive

evidence for the efficacy of LS (e.g., Cheung & Wong, 2014; Rzejak, 2019; Willems & van den Bossche, 2019) and struggles to use rigorous and comparable methods to evaluate LS outcomes (e.g., Cheung & Wong, 2014; Seleznyov, 2019). In addition, publications frequently include insufficient explanations of their LS intervention or the research methods that were employed (e.g., Baumfield et al., 2022; Larssen et al., 2018). Finally, the empirical research base offers several examples in which LS either failed to lead to any learning (e.g., Farhoush et al., 2017; Park, 2008) or was discontinued by schools (e.g., Brown et al., 2016; Dudley et al., 2019). These findings suggest that several questions remain open and that the advancement of Lewis et al.'s (2006) research needs remains critical to the field.

This dissertation therefore takes stock of the progress that has been made in the field of LS over the past 20 years. The overarching objective is to advance Lewis et al.'s (2006) research needs by means of three research studies and to identify future directions that can move the field forward. As this dissertation was conducted within the "Leistung macht Schule (LemaS)"-initiative (["Excellence in School Education"], BMBF & KMK, 2016), it also derives implications for research on LS in the German school context.

This dissertation is structured into three parts. The first part assesses the progress that has been made to date on each of the three research needs. To this end, two scoping reviews of the LS literature were conducted. The first review synthesizes all literature and systematic reviews of LS, while the second review focuses on models of teacher learning that have been either developed for or adapted to the context of LS. These reviews of the literature indicate moderate progress on Lewis et al.'s (2006) research needs and point towards four limitations that currently hinder improvement. These limitations are (1) the frequent lack of comparable and replicable descriptions of the LS intervention in publications, (2) the incoherent use or lack of theoretical frameworks to explain teacher learning through LS, (3) the inconsistent use of terminology and concepts, and (4) the lack of scientific rigor in research studies and of established ways or tools to measure the effectiveness of LS.

The second part of this dissertation presents three research studies that examine the extent and nature of these limitations. This dissertation puts an emphasis on the LS stages of observation and reflection, as these processes have been determined as mechanisms that can greatly facilitate teacher learning (e.g., Korthagen, 2016; Schön, 1995; van Es & Sherin, 2002). In the LS literature, these processes remain, however, particularly ambiguous and undertheorized (Larssen et al., 2018; Saito, 2012; Xu & Pedder, 2014).

The first study uses a mixed-method design to examine how four LS teams at German primary schools reflect together in regard to (1) their depth of discourse in terms of reflective stages, and (2) the respective trajectories through their reflective practice. In a first step, a theory-based definition for teachers' critical and collaborative reflection in the context of LS is established. The reflection process is then described in three stages that are derived from the ALACT model by Korthagen (1985) and Korthagen and Vasalos (2005); namely, describing observations, explaining and analyzing them, and finding solutions or courses of action. To examine how these refelction stages are enacted by LS teams, audio-recordings of four post-lesson discussions were collected at German primary schools. In line with Qualitative Content Analysis (e.g., Mayring, 2010; Schreier, 2012), audio-recordings were transcribed and analyzed using MaxQda (VERBI Software, 2019). For this purpose, a coding tool based on the ALACT model was developed and iteratively improved. Transcripts were coded by two coders and a satisfying inter-coder reliability of 0.82 % (Brennan's Kappa) was achieved. The schools' trajectories through their post-lesson discussion were analysed using micro-diachronic portraits created in MaxQda. Chi-square tests for independence were conducted to compare the frequencies of codes between schools. Findings indicate that the reflection processes of the four LS teams differed significantly and corroborate the view that phases of reflection are hard to distinguish from each other (Rodgers, 2002). The data indicates that the teams underwent mini-cycles of reflection (Slavit & Nelson, 2010), meaning that proposed solutions or insights were re-tested and adjusted by a further exploration of the topic. Teams struggled with certain aspects of their reflections, such as focusing their inquiry, prioritizing salient observations, and uncovering standard explanations. The findings imply that the collaborative and critical reflection in LS is a challenging process that needs to be routinized and practiced in order for teachers to be able to maximize their learning.

The second study, a systematic review, investigates previous findings that LS publications frequently lack key information concerning how the LS intervention was executed by teachers. Drawing on Moravcsik's (2020) three dimensions of research transparency, a coding protocol was established that details which information concerning the observation and reflection stages needs to be reported in LS research. The coding protocol, which was pre-registered on OSF (Kager et al., 2021), was used to assess 129 research articles on LS published in English between 2015 and 2020. The following research questions were examined: (1) How transparent are LS articles in reporting their observation and reflection stages?, and (2) which

theoretical frameworks are used in these studies to conceptualize the observation and reflection stages in LS? The findings confirm that the vast majority of articles underreport details such as how teachers enacted the classroom observation and the post-lesson discussion. In addition, only a minority of articles provided explicit definitions for these processes or grounded them in a specific theoretical framework. Several reasons for this lack of transparency, as well as consequences for the generation of knowledge in the field of LS, are discussed. Based on the findings, the study recommends a check list that can guide future empirical research in reporting their LS intervention.

The third study, a conceptual article, directly addresses Lewis et al.'s (2006) second critical research need and proposes a conceptual model for the field of LS. First, the scope and requirements of a model that can serve as a shared reference point to the field are determined by considering the research base on teacher learning, PD, and organizational psychology. Next, existing LS models are analyzed and several limitations are identified. These limitations are then addressed by proposing a new LS model that is designed along the IMOI structure (Mathieu et al., 2019) and combines concrete and theory-led inputs, processes, and outputs. The article specifies several ways in which the model can be applied by both researchers and educators.

The third and final part of this dissertation connects back to the limitations that hinder progress on the critical research needs in LS. Based on the findings of the research studies, new insights into the nature, cause, and extent of these limitations are discussed. The critical research needs by Lewis et al. (2006) are then updated and, looking forward, several strategies and practices to further advance these needs in the field as well as in the German context are deduced.

Table of Contents

List of Tables

Appendix

List of Figures

List of Abbreviations

Abbreviation	Meaning
ALACT model	**A**ction, **L**ooking back, **A**wareness of essential aspects, **C**reating alternative methods of Action, **T**rial
AR	Action Research
CLR	Collaborative Lesson Research
DBR	Design-based Research
IMOI	Input, mediating mechanisms, output, input
I-P-O	Input-Process-Output
LS	Lesson Study
PD	Professional Development
PDSA cycle	Plan-Do-See-Act cycle
WWC	What Works Clearinghouse

Note. Abbreviations appear in alphabetical order. Abbreviations used in the individual research studies are explained within the respective study.

Part I

Introduction and Theoretical Background

Part I: Introduction and Theoretical Background

Engaging in professional development (PD) is widely considered to be a key factor in teachers' ability to manage the challenges and expectations of their profession, and to continuously develop their expertise and competences (Darling-Hammond et al., 2009; Harland & Kinder, 2014). PD is commonly defined as any activities that support the professionalization of in-service teachers, including formal programs such as coaching and workshops, as well as informal activities, such as self-study (Coldwell, 2017). The main goal of PD is twofold: to enhance teachers' knowledge and improve their instructions, as well as to enhance student learning (Darling-Hammond et al., 2017; Guskey, 2002). The discourse on what constitutes effective PD has increased significantly in the past decades (e.g., Borko et al., 2010; Darling-Hammond et al., 2017; Hunzicker, 2011; Korthagen, 2016; Lipowsky & Rzejak, 2015) and gained relevance for several stakeholders, as scaling up PD programs requires a substantial commitment of resources (Guskey & Yoon, 2009). It has been difficult at times, however, to assess why PD approaches are successful in some, but not in other settings. Likewise, only some empirical studies have been able to reliably connect a PD intervention to an improvement in student achievements (Yoon et al., 2007). Reasons for this are, according to Guskey (2009), that conceptualizing and implementing a PD approach needs to go hand in hand with its critical evaluation, yet each of these steps presents methodological challenges.

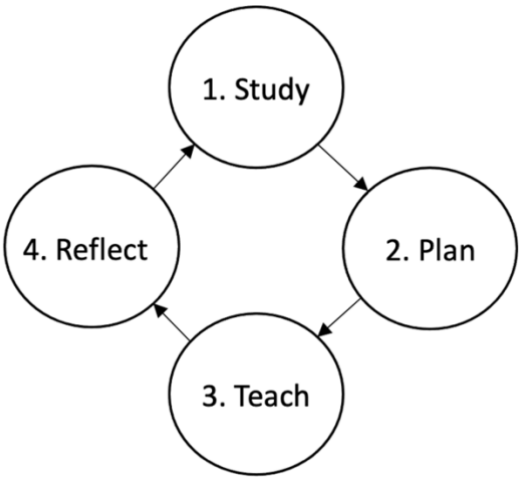
This dissertation addresses and investigates some of these challenges by critically reviewing a specific PD approach, namely Lesson Study (LS). The LS approach has its origins in Japan, but has received worldwide attention in the past three decades and rapidly spread around the globe (Yoshida et al., 2021). LS is a collaborative PD approach in which a team of teachers or pre-service teachers engages in iterative cycles of exploring their own instructions (Murata, 2011). According to Lewis et al. (2019), a LS cycle comprises a set of key stages: First, the LS team considers their classroom practice and studies the curriculum. The goal is for teachers to identify a topic or challenge that they want to explore together. After the team formulates a research question or lesson goal for their cycle, they collaboratively plan a lesson that addresses this question. Next, one team member delivers this lesson to a class in a so-called *research lesson*, while the other team members observe how students learn and react to the lesson. After the research lesson, the LS team comes together in the post-lesson discussion and analyzes their observations of student learning. The goal of this last step is to use the observations of student learning as evidence to collaboratively reflect on the lesson

plan, find answers to the research question, and generate solutions that each team member can then carry into their own practice.

In short, a LS cycle consists of the four stages illustrated in Figure 1: (1) Study, (2) Plan, (3) Teach, and (4) Reflect. Importantly, LS is an iterative and long-term approach, meaning that once a LS team concludes their cycle, it usually embarks on the next one. Teams are typically joined by so-called *knowledgeable others*—such as researchers, specialists, or teachers from other schools—, who offer their expertise to the team and/or facilitate the process (Lewis et al., 2019; Takahashi, 2014**).**

Figure 1

The Key Stages of the Lesson Study Process According to Lewis et al. (2019)



These four stages of LS are, on initial inspection, not new to teacher PD models traditionally employed in the US or Europe. Current classifications of PD frequently categorize LS as part of continuous improvement approaches (e.g., Dick et al., 2022; Lewis, 2015; Yurkofsky et al., 2020). This classification highlights that LS is an ongoing and sustained commitment to professionalization, addresses local problems and classroom needs, is driven by teachers, and aims to disseminate new insights beyond a single classroom or school (Yurkofsky et al., 2020). LS therefore shares common features with other collaboration-based models, such as teacher enquiry approaches, cycles of inquiry, professional learning communities, or data teams (Norwich, 2018; Yurkofsky et al., 2020). Helmke and Helmke (2019) and Rolff (2019), for instance, describe similar concepts employed in the German context that incorporate the

features of identifying certain goals or needs and subsequently planning and implementing an intervention that addresses these needs. Seleznyov (2018) argues, however, that LS can be clearly distinguished from other collaborative and continuous PD approaches by its explicit focus on teachers as researchers. While teachers do not conduct scientific research when engaging in LS, they do engage in processes commonly found in research: they identify a research question, study relevant materials and literature, formulate hypotheses or anticipate student learning, and then use their observations collected during the research lesson as evidence in order to arrive at data-led solutions and insights (Seleznyov, 2018).

For all these reasons, LS has been argued to incorporate many, if not most, of the features claimed as integral to effective PD models (Lewis et al., 2019; Murata, 2011). Darling-Hammond et al. (2017), synthesizing the empirical literature on the matter, identify seven critical features for effective PD: It is content focused, embedded in teachers' practice, engages teachers in collaboration, uses models of effective practice, provides external support, includes feedback and reflection, and is of sustained duration. These features, which several scholars converge on (e.g., Borko et al., 2010; Guskey & Yoon, 2009; Lipowsky & Rzejak, 2015), are all present in some form within the LS approach.

A steadily growing body of international and empirical research has investigated how these key features of effective PD translate to actual benefits of LS to teachers' on-going professionalization. Several studies, for instance, have connected LS to an increase in teachers' knowledge (e.g., Coenders & Verhoef, 2019; Knapp et al., 2011; Warwick et al., 2016), their awareness of student needs (e.g., Bruce et al., 2016; Cajkler et al., 2014; Dudley, 2013), and self-efficacy (e.g., Chong & Kong, 2012; Schipper et al., 2018). Other studies suggest that engaging in LS strengthens teachers' collaboration and supports the development of collaborative routines (e.g., Quaresma & da Ponte, 2019; Richit & da Ponte, 2019; Widjaja et al., 2017). In addition, some findings connect LS to a change in teachers' beliefs and attitudes (e.g., Hadfield & Jopling, 2016; Schipper et al., 2017). The empirical research base therefore suggests that LS has the potential to positively influence teachers' knowledge, skills, behavior, and beliefs.

These optimistic findings are, however, challenged by reports of less successful LS implementations and persisting misconceptions about its key activities (Chokshi & Fernandez, 2004; Fujii, 2014). While some studies describe that LS did not yield any measurable increase in teachers' knowledge (Brosnan, 2014; Callahan, 2019), others indicate that LS can also result

in “undesired” or “problematic” learning (Parks, 2008, p. 1214). The latter is discussed by Parks (2008), whose analysis of LS with pre-service teachers demonstrates that not all insights reached during LS are appropriate or substantiated. Instead, LS can also be used to confirm already held problematic beliefs and assumptions. This is in line with other studies that found that teachers might engage in LS in a superficial way (e.g., Bae et al., 2016; Canonigo, 2016; Mynott, 2019). In particular, empirical studies indicate that in order for LS to be effective, teachers need to possess the skills to systematically notice relevant classroom observations (Karlsen & Helgevold, 2019) and reflect critically in a group (Cammarata & Haley, 2018; Chikamori et al., 2013; Mynott, 2019).

The empirical research base on LS has therefore provided mixed results on the effectiveness of LS, despite the fact that LS incorporates most of the key features that have been claimed to make PD effective. This dilemma refers us back to the methodological challenges mentioned above that affect research into the effectiveness of any PD approach. First, it is inherently difficult to establish a measurable link between teachers’ participation in a PD and a change in teachers’ knowledge or students’ achievements (Guskey, 2021; Guskey & Yoon, 2009). This means that we need to develop rigorous methodologies informed by theories of learning that allow us to measure long-term outcomes and attribute these outcomes to a specific PD intervention. While progress has been made on this issue, scholars have yet to reach a consensus on the best way to conduct such a controlled evaluation in a school environment (Guskey, 2021), or for a concrete PD, such as LS (Willems & van den Bossche, 2019). The second challenge concerns the implementation fidelity of PD—meaning how close the actual implementation is to its intended implementation (Albers & Pattuwage, 2017). The four stages of LS, for instance, have been adapted to countless new contexts. As Hadfield and Jopling (2016) point out, research findings on these LS implementations are necessarily highly contextualized; what works in one setting might not yield the same results in another. This relates directly to the next challenge, namely the effort to scale-up evidence-based PD. As Cohen-Vogel et al. (2015) summarize, attempts at transferring evidence-based PD to new settings or to expand them, are not always successful. In the US context, for instance, Pogrow (2017) showed that a research-validated educational practice designed to improve children’s reading skills could not deliver positive results when transferred into schools. The reason why evidence-based innovations often struggle when scaled-up is that practice-based knowledge on how to implement these innovations for diverse populations

and in different settings is usually missing in research (Bryk, 2015; Cohen-Vogel et al., 2015). Century and Cassata (2016), who offer a perspective from implementation research, argue that investigating the impact of individual features of educational innovations could help to circumvent some of these challenges, yet others still apply even if interventions are deconstructed and evaluated step by step.

The overarching theme that guides this dissertation is therefore the question of how the international research community on LS has addressed the challenges of generating solid evidence for the effectiveness of LS over the past two decades, and to pinpoint open questions that can move the field forward. Given that LS is now being practiced on every continent and that the number of research studies on LS published each year has more than doubled since 2015 (Yoshida et al., 2021), this dissertation takes stock of the research field and reviews the progress that has been made. A number of systematic reviews on diverse aspects of LS research have already been conducted (e.g., Norwich et al., 2021; Seleznyov, 2019; Willems & van den Bossche, 2019). What is lacking, however, is an analysis of the research questions, or research needs, that initiated research on LS in the early 2000s (Lewis et al., 2006), and those research questions and needs that are relevant now to advance the field.

This dissertation is structured in three parts. The first part reviews the history of LS by discussing the critical research needs that spearheaded research into LS. It then assesses how the field has responded to these needs over the course of the past 15 to 20 years. Based on this evaluation, current research needs and questions are derived. In the second part, these research needs are investigated in three studies. The first study employs a mixed-method approach in order to theorize the reflection stage of LS and analyze how four LS teams reflect together. The second study develops a framework for transparent descriptions of the LS intervention in publications and systematically reviews how articles on LS report on the observation and reflection stages. Building on this systematic review, the third study seeks to bridge existing research on LS and theories on learning by proposing a conceptual framework for LS. The last part offers a discussion on the theoretical and practical contributions of these three scientific studies and identifies concrete solutions that can move the field forward, as well as critical research needs that remain open.

1. Then: Critical Research Needs in the Field of Lesson Study

This chapter provides a short overview of the internationalization of LS and discusses the beginning of the research field. The goal of this chapter is to better understand how and why LS was transferred from its original Japanese context to new settings around the globe, as well as what motivated and shaped the emergence of systematic research on LS.

Japanese LS started to attract international attention around three decades ago, when the first Trends in International Mathematics and Science Study (TIMSS) in 1995 revealed significant gaps in students' achievements between countries (Hiebert & Stigler, 2000). The results triggered the need to innovate classroom practices and teacher development systems in those countries that did not perform as expected (Xu & Pedder, 2014). US scholars in particular felt the pressure to advance their current teaching and professionalization methods and turned towards international perspectives on education for answers (Hervas, 2021b). The outstanding performance of Japanese students in the TIMSS was, as Hwang and Fwu (2011) explain, largely accredited to Japan's model of teacher education. According to Kawaguchi and Iwata (2021), the Japanese term *teacher education* is used to refer to the professionalization of teachers at any stage of their career, from initial teacher education to continuous PD for in-service teachers. This teacher education is heavily based on LS, meaning that pre-service teachers usually learn about LS and conduct LS in various forms during their teacher training and go on to participate in LS within and outside their schools throughout their career (Kawaguchi & Iwata, 2021). While LS in Japan can take on a variety of patterns, the common objective is always to improve teachers' instructions and students' achievements (Kawaguchi & Iwata, 2021).

The LS approach presented a fresh perspective to models of teachers' PD traditionally used in the West and resonated with the growing research base on which characteristics are key to enhancing teachers' professional knowledge (e.g., Borko et al., 2010; Darling-Hammond et al., 2017; Guskey & Yoon, 2009; Lipowsky & Rzejak, 2015). In particular, LS is not only embedded in teachers' daily practice, but gives teachers an active and central role in their own professionalization (Lewis, 2000). That is, Japanese teachers who participate in LS are generally in a position to ask for administrative support from their schools or the government, request that experts or educational specialists participate in their LS cycles, apply for additional funding to create new resources and learning materials, or publish and share their own learning in LS bulletins (Kim, 2021). This means that professionalization through LS in

Japan is largely teacher-driven and supported on a policy-level, which reduces conditions commonly identified as challenges to teachers' professional growth, such as time pressure, lack of financial or administrative support, lack of materials, or a mismatch between policies and the actual need for PD as experienced by the teachers (Darling-Hammond et al., 2017). For all these reasons, Japanese LS made an attractive PD model for other countries to adopt, which was especially true in the context of the "international atmosphere of dissatisfaction and disappointment with traditional teacher professional development practices" (Xu & Pedder, 2014, p. 30) that followed the TIMSS results.

Within a few years of its arrival to the US, LS spread to hundreds of schools across the country (Lewis et al., 2006) and the need for research on the effectiveness of LS in its new context became apparent. It has to be noted at this point, that research on LS in Japan was and remains scarce (Kim, 2021). Given that LS had already been used successfully for a century, researchers in Japan "did not feel empirical research was necessary to prove the approach's effect" (Kim, 2021, p. 24). This is reflected in the lack of scientific literature on LS prior to 2000 (Cheung & Wong, 2014). LS was thus not a new approach when it travelled to the US, yet research on LS, its core features, and its effectiveness was still in its infancy. As many US researchers and educators struggled to implement LS successfully in their contexts (Fernandez et al., 2003), it became clear that systematic research on the preconditions that would enable the successful implementation of Japanese LS in its host countries was necessary.

Research on policy borrowing in education—meaning when countries adopt reforms or innovations from another country in order to improve aspects of their own performance (Seleznyov et al., 2021)—has identified several challenges to the successful translation of policies to new contexts. According to Dolowitz (2009), borrowed policies might fail to convince the host country's stakeholders whose support is needed to provide crucial conditions for the implementation of the policy. Further, innovations might fail to be sustained beyond a project- or research-context, especially when the systemic structures to maintain practices are lacking (Dolowitz, 2009). Hadfield and Jopling (2016) add that innovations will likely have to be adapted to their new context, which requires multiple repetitions of trial and error. These adaptations need to take contextual and cultural factors into account (Grimsæth & Hallås, 2015), which might affect the degree of fidelity with which a policy can be transferred. All these challenges mean that many educational innovations fade away before

they can establish themselves (Burkhardt & Schoenfeld, 2003; Maddux, 2003; Maddux & Cummings, 2004; Tidd & Bessant, 2018).

This so-called *faddism* (Good et al., 1997; Lewis et al., 2006) of educational innovations typically follows a predictable cycle. According to Maddux and Cummings (2004), this cycle starts with an innovation being regarded as the new promising solution, followed by its' quick and often rushed implementation in schools. When this adoption does not yield the expected results, researchers, educators, and policymakers often experience disillusion, they abandon the innovation and subsequently turn to new promising ideas. This abandonment Is usually premature (Maddux & Cummings, 2004) and occurs before the innovations has been fully understood, or before a comprehensive research base has been developed (Burkhardt & Schoenfeld, 2003; Grimsæth & Hallås, 2015).

Concerns about this quick fading of reforms seem particularly present in the field of education around the turn of the century (e.g., Good et al., 1997; Maddux, 2003; Slavin, 1999). This coincides with the arrival of LS to the US, which was initiated by a few key articles written in the English language, most notably those by Stigler and Hiebert (1999), Yoshida (1999), Lewis (2000), and Fernandez and Yoshida (2004). These publications describe Japanese LS and discuss the first attempts of LS adaptations in the US. It was the impactful paper by Lewis et al. (2006), however, that first called for the development of a research base on LS—meaning a scientific knowledge base that goes beyond descriptions and reports of LS. Specifically, the authors identify three critical research needs for the then young but increasingly international field of LS (Table 1). These are (1) the need for a descriptive knowledge base of LS, (2) the need to explain how LS enables teacher learning, and (3) the need to test and refine how LS is implemented (Lewis et al., 2006). The article therefore marks an important turning point in the popularization of LS, as Lewis et al. (2006) formulate pressing research issues that need to be confronted in order to advance LS beyond the stage of "infatuation with an innovation" (Maddux, 2003, p. 122) and towards a well-researched and theoretically grounded PD model.

The first critical research need is the development of a strong descriptive knowledge base of LS (Lewis et al., 2006). In 2006, descriptions of Japanese LS were rare and according to the authors the understanding of LS in the US was based on only two documented examples (Lewis, 2002; Yoshida, 1999). The growing number of researchers and educators

across the US that were adapting LS to their contexts therefore had little resources to draw on. In their article, Lewis et al. (2006) note that Japanese publications on LS would be of "great practical and theoretical interest to U.S. educators" (p. 4), especially in order to keep the key features of LS intact despite its transfer to new settings. In order for LS to survive its first hype in the US, Lewis et al. (2006) therefore argue for the need of rich descriptions and examples of LS that can serve as models for others.

Table 1

Overview of Critical Research Needs in LS: Then

	Research Need 1	Research Need 2	Research Need 3
Then	**Expansion of the Descriptive Knowledge Base of Japanese and U.S. Lesson Study**	**Explication of the Innovation Mechanism**	**Design-Based Research Cycles**
Research needs in LS according to Lewis et al. (2006)	to expand the descriptive knowledge base on LS in an effort to describe LS's characteristics and determine adaptations pertinent to LS's implementation in US settings	to investigate the mechanisms through which teachers learn in LS and develop a model that represents these mechanisms as well as LS surface features	to use design-based research cycles to improve LS adaptations and support theory-building

The second critical research need identified by Lewis et al. (2006) is the investigation of LS's innovation mechanism. By *mechanism*, they refer to underlying processes that enable teacher learning. LS consists of four stages, or surface features, but professional learning is theorized to result from the underlying processes that teachers engage in when implementing these surface features (Boylan et al., 2018). Several models that aim to capture the exact nature of teacher learning have been developed in the field of education (e.g., Clarke & Hollingsworth, 2002; Desimone, 2009; Guskey, 2002). Lewis et al. (2006) also offer their own framework: LS allows teachers to engage with learning resources, increase their knowledge, and consider their personal motivation to teach, which then leads to the improvement of instruction. Putting this framework forward for discussion, they emphasize the need for a model of LS that captures not only its surface features, but also its underlying mechanisms.

The third critical research need identified by Lewis et al. (2006) is theory-building through design-based research cycles. Specifically, they argue that researchers need to "progressively hone" (p. 5) LS by continuously connecting practical LS experience to theoretical views. They

suggest design-based research cycles as an approach to collect data on LS implementation and then analyze this data in order to advance the field's theoretical understanding of LS.

Lewis et al. (2006) conclude their article with a critical comment on the problem of faddism. One of the prevalent explanations at the time for why educational innovations tended to wash out quickly was that these innovations had been insufficiently researched and could not be considered evidence-based (Lewis et al., 2006). To counter the fading of educational innovations, policy makers therefore aimed to only adopt and scale-up innovations that had been proven effective through, for example, randomized-controlled trials (Bryk, 2015; Lewis et al., 2006). Lewis et al. (2006) criticize this approach by noting that conducting experimental research on "immature versions of lesson study" (p. 10) was likely to yield unsatisfactory results, leading to the disillusionment and the abandonment of LS. In order to break the cycle of adopting and abandoning new innovations, researchers should instead invest time into investigating how LS could work in the US context, before putting it to the test by means of controlled experimental research.

Lewis et al. (2006) therefore problematize how research on educational policies was traditionally conducted and point out the challenges that the field faced at the time. Specifically, that experimental research tends to be costly, slow, and difficult to transfer to new contexts (Bryk, 2015). Other types of research, such as qualitative and small-scale studies, tend to be less costly, but are not always considered "credible" (Burkhardt & Schoenfeld, 2003, p. 3), as they often lack scientific rigor and cannot provide causal relationships between interventions and practice. In order to circumvent these limitations, Lewis et al. (2006) argue for the "local proof route" (p. 7): namely, to implement and study innovations locally and increase its effectiveness through repeated cycles of improvement. These innovations are then spread organically or planned, but with flexible fidelity that allows the innovation to be adapted to new contexts. This approach stands in contrast to the "general proof route" (Lewis et al., 2006, p. 7), which requires the controlled study of innovations and its subsequent spread with high fidelity.

In this chapter, we have considered the history and first steps in the internationalization of LS. We then identified the article by Lewis et al. (2006) as a milestone in the emergence of the research field of LS. By pointing towards the difficulty of sustaining borrowed policies and educational innovations, Lewis et al. (2006) made a compelling case for the need of systematic research that focuses on describing LS, theorizing LS, and adapting and improving LS over

continuous cycles. In the next chapter, we review the subsequent development of the field and assess the progress that has been made to date on these research needs.

2. Now: Taking Stock of Current Research Needs

The field of LS has grown significantly since the publication of Lewis et al.'s (2006) article on critical research needs in LS. This chapter focuses on assessing whether Lewis et al.'s (2006) research needs have been met, whether they should still be the priority in the field, and which questions remain open.

For this purpose, two scoping reviews of the LS literature were conducted. The goal of these reviews was threefold: (1) to synthesize the available knowledge and findings that have been generated in the field, (2) to clarify the progress that has been achieved in regard to the critical research needs identified by Lewis et al. (2006), and (3) to pinpoint knowledge gaps as well as remaining or new research needs. Following Munn et al.'s (2018) recommendations for conducting a scoping review, several databases (SCOPUS, ERIC, PsychInfo, Academic Search Premier, Bibliography of Asian Studies, JTSOR, and ProQuest) were searched with pre-defined search strings. In order to cover all three critical research needs identified by Lewis et al. (2006), the first scoping review synthesized all reviews that have been conducted on LS, while the second explored models of how teachers learn through LS.

For the first review, all peer-reviewed literature reviews and systematic reviews on LS published in English or German until November 2022 were determined (search strings: "lesson study" AND ("review" OR "systematic review" OR "literature review" OR "meta-analysis" OR "synthesis")). The search produced 21 relevant reviews, one of them written in German (Rzejak, 2019). Additional reviews were excluded due to them primarily targeting the typology of LS research (Chen & Zhang, 2019; Saito et al., 2020), reviewing LS projects conducted by a specific university (Soto Gómez et al., 2019), being beind a pay wall (Burrows, 2022), or not being available in English or German (Murase, 2007; Ono, 2009). It is recognized that the focus on publications written in English or German presents a delimitation.

The main findings and implications of each review are provided in Table A1 (see Appendix). Figure 2 provides a visual overview of the included reviews grouped into clusters according to their respective research aims. The illustration indicates that five reviews have addressed the available evidence for the effectiveness of LS (Cheung & Wong, 2014; Rzejak, 2019; Seleznyov, 2019; Willems & van den Bossche, 2019), with Kanellopoulou and Darra (2019) focusing on LS in higher education. Several reviews have focused on describing the state-of-the art of the field (Huang & Shimizu, 2016; Saito, 2012) and aspects of the field's growth and geographical spread (Hervas, 2021b; Xu & Pedder, 2014; Yoshida et al., 2021). More recently, reviews have

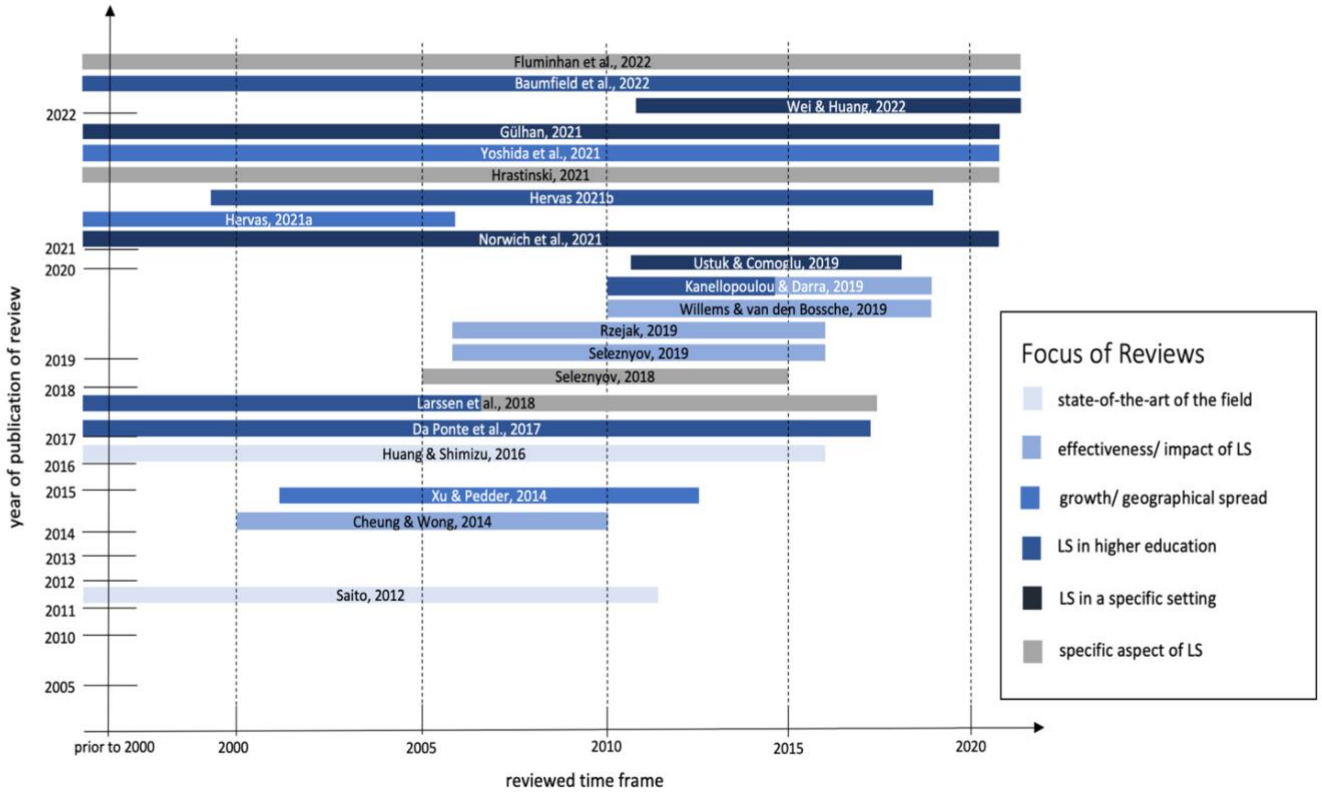
targeted LS in the setting of higher education (Baumfield et al., 2022; da Ponte, 2017; Hervas, 2021a; Larssen et al., 2018). Four reviews have investigated LS in specific contexts, such as a region or country (Gülhan, 2021; Wei & Huang, 2022), a subject matter (Uştuk & Çomoğlu, 2019), or inclusive education (Norwich et al., 2021). Four reviews have examined specific aspects of how LS is implemented, namely how teacher learning and the observation of student learning is conceptualized in LS with pre-service teachers (Larssen et al., 2018), how the steps of Japanese LS are implemented internationally (Seleznyov, 2018), and the use of digital tools (Hrastinski, 2021). Lastly, the review by Fluminhan et al. (2022) reviewed the relationship between LS and teachers' self-efficacy.

The second scoping review concentrated on synthesizing models and conceptual frameworks of how teachers learn through LS (search strings: "lesson study" AND ("model" OR "conceptual" OR "conceptual framework" OR "conceptual model" OR "theoretical", "theoretical model" OR "process" OR "learning" OR "teacher learning" OR "mechanism"). The search for eligible peer-reviewed articles written in English proved challenging. Several articles, for instance, have developed or adapted models of teacher learning as part of their theoretical background, but do not include relevant keywords in their abstract. Some of these articles were identified through backtracking sources referenced in other articles. In addition, some relevant articles, such as Murata et al. (2004), were not publicly accessible. The current synthesis includes 18 articles, but there are likely additional eligible studies that could not be identified or accessed during the scoping of the literature. These 18 articles are listed in Table 2 (see section 2.2.2.).

In the following sections, the progress on each of the three critical research needs identified by Lewis et al. (2006) is assessed based on these two scoping reviews of the current LS literature.

Figure 2

Overview of Literature and Systematic Reviews Grouped into Clusters According to Their Respective Research Aims



Note. The focus "LS in a specific setting" refers, for instance, to LS within a specific country or a specific school setting.

2.1. Research Need 1: Development of a Knowledge Base on Lesson Study

The first research need identified by Lewis et al. (2006) was to produce descriptive resources, both practical and theoretical, that educators and researchers could draw on when implementing LS outside of Japan. Lewis et al. (2006) specifically refer to a *descriptive* knowledge base, but do not further define this term. This dissertation adopts a comprehensive perspective—meaning that it examines the general professional knowledge base of LS**.** The next section is therefore concerned with deriving a working definition for the term *professional knowledge base*.

2.1.1. What Constitutes a Professional Knowledge Base?

In order to explore what a professional knowledge base for educational research, and in particular for the field of LS, looks like, we need to consider four aspects: who contributes the knowledge, how is it generated, how is it systematized, and how do scholars and educators engage with it?

The first aspect addresses the fact that the field of education, much like the fields of medicine or law (Davidoff et al., 2015; Hiebert et al., 2002), needs to be informed by both theory and practice, and ideas from each domain need to be transferred to and tested within the other (Burkhardt & Schoenfeld, 2003). This bridging of theory and practice in education is not always successful (Bakkenes et al., 2010; Korthagen, 2016). As Cohen-Vogel et al. (2015) note, theoretical ideas about effective practice often fail in the classroom and educational research is not always accepted or used by educators. Practical knowledge and ideas developed by educators in the field, on the other hand, might be viewed as not trustworthy or widely applicable by scholars (Cohen-Vogel et al., 2015). In order for the field to be successful in spite of such tensions, it is crucial to find ways to transfer insights from large-scale research trials to the classroom, and to extrapolate useful and generalizable insights from concrete and contextualized field-research (Burkhardt & Schoenfeld, 2003). A sound knowledge base of LS therefore requires a strong partnership between research and practice.

The second aspect concerns how knowledge is generated. As Guskey and Yoon (2009) argue, a knowledge base on the effectiveness of PD programs needs to be first and foremost trustworthy, meaning that it meets scientific standards and that findings are verifiable and replicable. This demands that research studies employ methodologically sound designs, which are rigorously documented in a standardized and accessible way. For these reasons,

randomized controlled trials and quasi-experimental designs are traditionally considered the gold standard in producing credible insights (Guskey & Yoon, 2009). However, small-scale research on how the effectiveness of practices can be improved in local contexts is becoming increasingly important in the field (Bryk, 2015; Cohen-Vogel et al., 2015). A knowledge base of LS therefore needs to integrate the result of a variety of research approaches, including experimental and controlled approaches as well as practice-informed and descriptive methods.

Building on this knowledge generation, the third aspect addresses how a knowledge base is systematized. The process of organizing knowledge starts by embedding findings into a theoretical framework that permits researchers to talk about concrete insights on an abstract level (Hiebert et al., 2002). As Wang et al. (2020) argue, the theoretical underpinnings that shape a research field and offer a shared frame of analysis are crucial in order for research output to be understandable and applicable across the field. A strong theorization thus enables the systematic generalization of knowledge and allows researchers and educators to "move what was learned in one context or classroom into another" (Hiebert et al., 2002, p. 8).

This brings us to the last aspect, namely how researchers, educators, and policy makers engage with insights. A knowledge base needs to be dynamic and evolving, as "archived research knowledge" (Hiebert et al., 2002, p. 3) usually does not impact teachers' instructions in the classroom. This means that knowledge needs to be recorded, stored, and shared in a way that it is accessible to everyone with an interest to use, test, and build upon this knowledge. This can be challenging for a research field like LS that is international, yet always influenced by local and national contexts (Hadfield & Jopling, 2016). As Lewis et al. (2006) note, the insights collected over a century by Japanese educators and researchers were largely undocumented when LS travelled to the US, and the few scientific publications that existed were almost exclusively written in Japanese. In order for an international research field to accumulate a rich descriptive knowledge base, differences in language and research paradigms need to be tackled and resolved. These challenges also affect knowledge sharing on the national and local level. In Japan, the transmission of LS occurs on an institutional level, as knowledge about how to conduct LS is "reproduced through the Lesson Study-based teacher education system" (Kim, 2021, p. 21). This means that Japanese teachers share a "common conceptual grid" (Kim, 2021, p. 21) of LS that enables them to talk about it in established terminology. Outside of Japan, this systemic transmission of LS has yet to be

established. Researchers and scholars around the world consequently develop their own terminology and frameworks that best fit their contexts. A shared theoretical and conceptual framework could help the field of LS to remain coherent in spite these challenges.

The professional knowledge base on LS is therefore the product of both researchers and educators, it is generated by means of rigorous, diverse, and transparent research methodologies, it is systematized in a coherent and intelligible way, and it is accessible to and useable by a wide audience.

2.1.2. Assessing the Current Knowledge Base on Lesson Study

We will now consider the four aspects discussed above in order to assess the current knowledge base of LS.

Who is contributing to the LS knowledge base? The synthesis of literature and systematic reviews of LS suggests that the topic of how researchers and educators interact and jointly generate knowledge on LS has not yet been thoroughly investigated or reviewed. There are some indications, however, concerning who the stakeholders in LS are, as well as about the components of geography and educational settings. Concerning the stakeholders, the synthesis of reviews identified several roles that are critical to LS, namely teachers, school leaders and administrators, text book publishers, policymakers, and researchers (Huang & Shimizu, 2016; Saito, 2012). According to Huang and Shimizu's (2016) review, teachers in Japan and China often take on a leadership role in LS and actively participate in generating research findings. Beyond Japan and China, LS seems to be predominantly initiated and led by researchers, and is therefore predominantly connected to specific projects (Huang & Shimizu, 2016). Wei and Huang (2022) comment that the partnership between researchers and educators is often characterized by hierarchical structures. Teachers should, however, be treated as "key stakeholders" (Wei & Huang, 2022, p. 150) and actively participate in research, instead of merely being the object of research. The synthesis of the literature offers little information on how or whether this relationship has been investigated in the literature so far.

Aspects of geography and educational settings are better documented in the LS literature. The mapping review by Yoshida et al. (2021), for instance, demonstrates that over 40 countries across all continents are currently publishing research on LS. The majority of this research originates from Asia (Yoshida et al., 2021), yet several reviews identify the US as the country that produces the most research studies (Seleznyov et al., 2021; Xu & Pedder, 2014), especially

in the setting of higher education (Baumfield et al., 2022; da Ponte, 2017; Hervas, 2021a). There is a strong research output from European countries, and insights from South America and Africa are increasing, but remain relatively scarce (Yoshida et al., 2021). The synthesis of the reviews further shows that LS is being practiced across all levels of education (Yoshida et al., 2021). While the majority of research targets LS with in-service teachers (Larssen et al., 2018), Figure 2 suggests a growing interest into LS in the context of higher education. This means that the knowledge base on LS includes a variety of national and local perspectives and addresses several educational settings and adaptations of LS.

How is LS knowledge generated? The synthesis of reviews indicates that researchers converge on the fact that the field currently lacks trustworthy and rigorous evidence for the effectiveness of LS. Several reviews report that the majority of research on LS is qualitative, small-scale, and employs some kind of explorative design (Gülhan, 2021; Norwich et al., 2021; Rzejak, 2019; Seleznyov, 2019; Xu & Pedder, 2014). These studies offer weak explanatory power (Xu & Pedder, 2014), especially since several reviews report the use of different and sometimes inappropriate outcome measures (Cheung & Wong, 2014; da Ponte, 2017; Willems & van den Bossche, 2019). Likewise, some reviews found that articles lack information that is crucial to understand their research methodology or their LS intervention (Baumfield et al., 2022; Cheung & Wong, 2014; Larssen et al., 2018).

Four reviews focused exclusively on synthesizing the best evidence in the field on the effectiveness of LS. Both Cheung and Wong (2014) and Rzejak (2019) identified nine studies that used controlled designs, while Willems and van den Bossche (2019) included only five studies. All three reviews conclude, however, that the evidence of the effectiveness of LS cannot yet be confidently established. Seleznyov (2019) links this lack of trustworthiness to the methodologies employed in the reviewed studies. Using the Maryland Scientific Method Scale (Waights, 2014), her review demonstrates that only two out of 56 reviewed studies scored the maximum points on the scale; namely the studies by Lewis and Perry (2017) conducted in the US and by Murphy et al. (2017) conducted in the UK. Both studies employed a randomized-controlled design and assigned teachers to two conditions (i.e., group with LS intervention, group without LS intervention). The study by Lewis and Perry (2017) included a third condition: teachers engaged in LS and were additionally provided a mathematics resource kit. Interestingly, both studies failed to connect the LS condition to improved post-test scores. Lewis and Perry (2017), however, were able to show that the teachers and

students assigned to the third condition (i.e., LS intervention and a resource kit) achieved significantly higher scores after the intervention than the other two groups.

How is LS knowledge systematized? Several reviews emphasize the lack of theoretical frameworks in LS research, with Yoshida et al. (2021) noting that studies that theorize LS predominantly originate from Europe. Da Ponte (2017) found that some of the reviewed studies on LS in initial teacher education did not connect their findings to any theoretical framework. Larssen et al. (2018), also reviewing LS in initial teacher education, confirmed this lack of coherent theorization. The synthesis of reviews further suggests that particular areas are undertheorized in the LS literature, namely how teachers learn (Larssen et al., 2018; Xu & Pedder, 2014), and how they observe and reflect together (Kanellopoulou & Darra, 2019; Larssen et al., 2018; Saito, 2012). Saito (2012), in his early review, called on the research field to investigate the stages of observation and reflection. Findings from the second scoping review indicate that a number of models that aim to fill this gap have been formulated. The summary of these models in Table 2, however, suggests a lack of coherence concerning this theorization (for further discussion and Table 2 see section 2.2).

How does the LS community engage with the LS knowledge base? The last aspect concerns how the field is engaging with its knowledge base. The synthesis suggests that a lack of shared outcomes measures and precise descriptions of research methodologies currently challenge the replicability of studies (Cheung & Wong, 2014; Willems & van den Bossche, 2019). This is reflected in Cheung and Wong's (2014) comment that the field cannot conduct a meta-analysis unless it adopts a higher scientific standard in its publications. Concerning the implementation of LS in practice, the synthesis suggests that the field has produced an abundance of LS descriptions in a variety of contexts and countries that can serve as guidance to other researchers and educators. These descriptions, however, are also limited in their usefulness as they routinely lack crucial information (Baumfield et al., 2022; Larssen et al., 2018).

2.1.3. Limitations to the Current Knowledge Base

The assessment of the current knowledge base on LS indicates that Lewis et al.'s (2006) first critical research need has been answered only in parts. While the field has generated an impressive body of research and descriptive knowledge on LS, this knowledge is subject to several limitations.

First, the scoping of the literature indicates that contributions to the knowledge base on LS are international, but predominantly stem from the US and Asian countries. As Saito (2012) and Rzejak (2019) note, research on LS in Japan, the country of LS's origin, is still rare. We further saw that LS outside of Japan is frequently initiated by researchers and connected to a specific project (Wei & Huang, 2022). Yet, there seems to be little research or documentation on how the relationship between practice and science is viewed, whether educators are involved in conducting research on LS, and if yes, in what capacity. It has been acknowledged in the LS literature that LS research does not always pursue the same goals. As Stigler and Hiebert (2016) point out, some refer to LS as a research methodology, with teachers being active participants not only in LS, but in the research process. Others view LS research more narrowly as the assessment of a PD approach and conduct research in order to evaluate whether and what teachers have learned through participating in LS (Stigler & Hiebert, 2016). It seems, however, that this topic and the potential tensions connected to it have not yet gained significant attention in the current LS literature.

Second, the synthesis of reviews found that methodological approaches and outcome measures differ between studies, and that descriptions of both the research approach and the LS intervention are frequently incomplete in published research. This means that, currently, the field's ability to engage with and build on its own research and findings is limited. These problems are not confined to the field of LS, but mirror challenges that affect the evaluation of PD in general. As Guskey and Yoon (2009) criticize, only few studies that assess the effectiveness of PD meet strict scientific criteria, and the information about the PD intervention provided in publications tends to be incomplete and "far from perfect" (p. 496).

In order to make progress on these issues, it seems that Lewis et al.'s (2006) research need to build a descriptive knowledge base on LS is still of critical importance to the field. However, the research need should be updated and defined more closely. As we have already accumulated an international and descriptive knowledge base, we should now focus on building an interdisciplinary base of research that (1) uses theory-based and rigorous approaches to study teacher learning through LS and (2) provides transparent and complete descriptions of their research methodologies as well as of their LS intervention. The following open questions could support the building of this knowledge base: (1) Which features are currently underdescribed in LS research and how can we facilitate complete descriptions of these features in research?, (2) What are the differences between research on LS as a PD

approach and research on LS as a research methodology?, and (3) Are there methodological approaches from other research fields that could improve the way LS research is conducted?

2.2. Research Need 2: Explication of Lesson Study's Mechanisms

The second research need identified by Lewis et al. (2006) was to theorize and illustrate the processes that make LS effective and establish relations between certain activities (i.e., planning a lesson) and their outcomes (i.e., improved instructions). This section first discusses current theoretical perspectives on teacher learning and related concepts, and then narrows in on the use of these perspectives in the field of LS. The terms *theoretical* and *conceptual framework* tend to be used interchangeably in the literature. In this dissertation, these terms refer to any account that hypothesizes that two or more events are linked and affect each other in a substantial way (Davidoff et al., 2015). The term *model*, on the other hand, is used to refer to an illustration or map of such a theoretical account, that aims to depict a complex phenomenon in a systematic and abstract way (Davidoff et al., 2015).

2.2.1. Models and Theories of Teacher Learning

Several models and theories that aim to conceptualize teachers' learning processes have been proposed. Initially, teacher learning was predominantly conceptualized as a process of cause and effect. Guskey (2002) for instance, assumes that PD leads to changes in teachers' classroom practices, which leads to changes in students' learning outcomes, which then leads to changes in teachers' beliefs about their practice. This model conceptualizes that teachers first need to observe positive results in the classroom, before they adjust their beliefs and attitudes. Similarly, Desimone (2009) theorizes that teachers who engage in PD increase their knowledge and change their beliefs, which then leads to improved instructions and student learning. While the order of events differs, both models assume a linear relationship between the start and end point and offer limited information on the specific processes that drive this trajectory.

In current literature, a more dynamic approach is taken, where it is acknowledged that learning is a complex and individual process (Kennedy, 2016; Korthagen, 2016; Opfer & Pedder, 2011). These perspectives usually build on Clarke and Hollingsworth's (2002) interconnected model of professional growth, which assumes that teachers learn by engaging in iterative cycles of practice and critical reflection. The model suggests that learning can follow not one, but several interconnected pathways, and that these pathways differ from

teacher to teacher. It further shifts the focus away from linear events and towards iterative processes that mediate change, namely enactment and reflection.

Clarke and Hollingsworth's (2002) conceptualization of learning as a complex process driven by reflection connects to several theoretical perspectives on learning, such as transformative learning theories (Mezirow, 1990, 2000), the reflection-on-action perspective (Schön, 1983, 1995), or the conceptual change theory (Limon, 2001; Posner et al., 1982; Vosniadou et al., 2020). In general, learning can be regarded as the adding of previously missing knowledge, the gap filling of incomplete knowledge, or the change of existing knowledge (Chi, 2008). The latter kind of learning refers to the scenario in which a person receives new information that conflicts with their current knowledge and beliefs, and thus requires the reexamination and adjustment of these beliefs (Chi, 2008; Posner et al., 1982). This process, also known as the conceptual change theory, is particularly useful to conceptualize how teachers continue their professional learning throughout their education and career (Vosniadou et al., 2020). As Korthagen (2016) notes, teachers hold or acquire a host of often "limiting beliefs" (p. 400) that guide their classroom instructions. PD that engages teachers in critical reflection and exposes them to new ideas can stimulate the reexamination of such beliefs.

Many researchers posit that the central mechanism that induces this conceptual change is *cognitive dissonance* (Festinger, 1962; Harmon-Jones & Mills, 2019; Posner et al., 1982). Cognitive dissonance—also frequently referred to as *cognitive conflict*—describes the state of dissatisfaction or conflict between existing and new concepts, and the human need to resolve such conflicts (Kang et al., 2004). Research in education and social psychology has produced mixed results on whether cognitive conflict does in fact lead to measurable learning (e.g., Hinojosa et al., 2017; Vosniadou et al., 2020). Other research has focused on identifying ways in which people react to cognitive conflict. While some people succeed in reexamining their prior beliefs, others tend to ignore and avoid conflicting information, or simply ignore aspects that do not fit their already held views (Johnson & Johnson, 2009; Murray, 1983; Shu et al., 2011). Theories on conceptual change and reflection therefore emphasize individual cognitive mechanisms through which a person acquires new knowledge and, importantly, reexamines and adjusts existing knowledge.

Social learning theory (Bandura, 1969; Vygotsky, 1986) takes a slightly different perspective on learning by highlighting the aspects of collaboration, language, and situational

context. Specifically, this theory assumes that knowledge is co-constructed through talk and interaction. Language is therefore considered the primary tool that drives knowledge building and also shapes how knowledge is structured in the mind (Vygotsky, 1986). This theoretical account of learning is still widely used in the fields of education, linguistics, and psychology, and has also been translated to methodologies that aim to examine this learning process. These include, among others, cultural historical activity theory (e.g., Igira & Gregory, 2009; Sannino & Engeström, 2018) and sociocultural discourse analysis (Mercer, 2004).

This short review of theoretical approaches to teacher learning highlights that we can distinguish between models that aim to illustrate the process of professional learning (e.g., Clarke & Hollingsworth, 2002; Desimone, 2009; Guskey, 2002), and perspectives that theorize how aspects of these models interact which each other (conceptual change theory, sociocultural theory). There is one thing common to all these approaches: the orientation towards an outcome, namely the acquisition of new knowledge. Yet, the acquisition of knowledge on part of the teacher is not the final, but only an interim outcome on the trajectory to ultimately increase students' knowledge. This means that models of teacher learning need to conceptualize several outcomes over time that build on each other. While models such as those by Guskey (2002), Desimone (2009), and Clarke and Hollingsworth (2002) are oriented towards these outcomes, they are not designed to conceptualize how these outcomes develop over time, or how this development can be measured scientifically. Boylan et al. (2018) therefore distinguish between general models of professional learning and models that categorize outcomes of professional learning and therefore allow to evaluate the effectiveness of a certain PD approach.

An early but highly influential example of the latter type of model are the five levels of PD by Guskey (2000). Guskey conceptualizes hierarchical levels of outcomes that range from immediate to long-term outcomes. The first and immediate level concerns the participants' reaction to the PD, which can be evaluated through questionnaires that include rating-scale items and open-ended responses. The second level concerns whether participants learned something from the PD. Depending on the type of PD, this increase in knowledge might be assessed by means of a test or an oral or written reflection, but also through a simulation or demonstration in the classroom. The third level concerns the gradual impact a PD has on the organizational structures and routines of the school it is embedded in. This includes administrative support to implement changes, the sharing of knowledge across or beyond a

school, and whether resources necessary to maintain changes are made available to teachers. The success of this level can be measured by means of questionnaires and interviews, but also school records, minutes of meetings, or similar protocols. The fourth level concerns whether participants succeed in transferring their new knowledge into their daily practice. This long-term outcome aims to evaluate whether teachers can use their new skills or insights in the classroom and manage challenges that might arise in the process. This level can, again, be measured by means of questionnaires and interviews, but also by observation and video tapes. The last level concerns change in students' learning outcomes, such as their achievements, performance, well-being, or self-efficacy. This level can be evaluated by looking at student or school records, or conducting questionnaires or interviews.

According to Guskey (2000), only the evaluation of all these levels can demonstrate whether a PD had any lasting impact. He further argues that PD should be designed backwards: the specific student outcomes should be the first consideration when developing a PD intervention aimed to achieve these outcomes.

Guskey's (2000) approach to classify PD outcomes is still widely used in the field of education and has informed the past two decades of empirical research on effective PD. The conceptualization of hierarchical levels of outcomes has also informed the development of other models. Lipowsky and Rzejak (2015), for instance, propose a so-called offer-and-use model for research on teachers' PD that incorporates Guskey's levels. Specifically, their model aims to systematize components of the PD (offer) and their effect on participants (use). The authors conceptualize a range of factors that influence the participants' use: the participant's characteristics, the facilitator's characteristics, the school context, the quality and quantity of learning opportunities, as well as the participants' perception of these opportunities. These factors and their interplay lead—through a so-called *transfer-process*—to four levels of outcomes: participants' satisfaction with the PD, the enhancement of participants' knowledge and their instructions, as well as the enhancement of student performance (Lipowsky & Rzejak, 2015, p. 30). Lipowsky and Rzejak's (2015) model is widely used in research on PD, especially within the German context.

If we take a step back and consider the structure of the models proposed by Guskey (2000) and Lipowsky and Rzejak (2015), we can see that Guskey focuses solely on outcome levels. Lipowsky and Rzejak, on the other hand, integrate these levels in a wider conceptualization of PD, similar to earlier models, such as the CIPP model by Stufflebeam (i.e., context, input,

process, and product; Stufflebeam, 1983). By doing so, they systematize how certain input factors trigger a process that yields certain outputs. This structure is commonly referred to as I-P-O structure (i.e., input-process-output; Driskell et al., 2018; Ilgen et al., 2005) and has been used as a heuristic to explain a variety of processes, such as team effectiveness, in the fields of education, management, and psychology (e.g., Marks et al., 2008; Mathieu et al., 2019). The advantage of models that follow the I-P-O structure is that they can potentially tie together the two ideas that, according to Boylan et al. (2018), tend to remain separated in the literature: how teachers learn and how this learning can be conceptualized in concrete outcomes.

This section has provided a brief insight into models and theories of how teachers learn. The next section reviews the use of these models and theories in the field of LS.

2.2.2. The Use of Models and Theories in the Field of Lesson Study

The most commonly used model in the LS literature is the circular depiction of the four LS stages (see Figure 1). This model has been employed and adapted countless times in the LS literature (e.g., Dick et al., 2022; Dudley, 2013; Joubert et al., 2020; Lewis, 2009; Moss et al., 2015). While it illustrates the activities that teachers engage in during a LS cycle, it does not conceptualize teacher learning. The present analysis concentrates on models that go beyond this depiction and incorporate some or all of the dimensions of input, process, and output. The basis of this discussion is Table 2, which lists the 18 articles that were identified during the second scoping review of the literature. Specifically, the table details whether and how a model or framework specifies the dimensions of input, process, output, or area of impact. The table further indicates whether the model or framework is based on a specific existing approach and whether they were developed specifically for a certain piece of LS research.

To begin with, Table 2 distinguishes between the aspects of *input* and *structural features of LS*. According to definition, *input* refers to “antecedent factors that enable and constrain [team] members’ interactions” (Mathieu et al., 2019, p. 18). The dimension of input therefore comprises structural features of a PD intervention, as they are predefined and determine the quality and quantity of learning moments. The decision was made to list the two aspects of *input* and *structural features of LS* separately in Table 2, in order to make visible the predominant lack of input factors that extend beyond structural features. As Table 2 demonstrates, more than half of all models limit inputs to LS activities. Only eight models

include additional input factors, such as participants' characteristics, the school context, support, or resources. These additional input factors differ between models, and also remain rather abstract in some of them.

Table 2

Overview of Models of How Teachers Learn Through Lesson Study

study	year	input	structural features of LS	process	area of impact	output	character-istics	informed by	project specific	page
Lewis et al.	2006		observable features of LS (detailed description provided in model)	strengthening of pathways (intervening changes)	intervening changes in teachers' knowledge, teachers' commitment, and community, learning resources	improvement of instruction	I-P-O	e.g., Lewis et al. (2005)		p. 5
Lewis	2009		LS cycle (planning, Curriculum Study, research lesson, data collection, discussion, revision, etc.)		knowledge, interpersonal relations, personal dispositions	instructional improvement	I-P-O	Lewis et al. (2006)		p. 100
Lewis et al.	2009		features of LS (investigation, planning, research lesson, reflection)	thinking becomes visible, community norms, tools, identity, and participation develop	intervening changes in teachers' knowledge and beliefs, teachers' professional community, and teaching-learning resources	two sequential outcomes: instructional improvement leading to student learning	I-P-O	Lewis et al. (2006); cognitive theories of teacher learning and situated learning theories		p. 287
Murata	2011		LS Activities (detailed description provided in model)		teacher learning: knowledge base, commitment to community, resources	improved classroom practice	I-P-O with feedback loop	Murata et al. (2004)		p. 6
Ylonen & Norwich "The Lesson Study process model"	2013	knowledge base of teachers, Whole school conditions/ Senior leadership support	planning, lesson, observation, review, create artefact to share learning	observation, review		create artefact to share learning	includes some features of I-P-O, illustrates serval cycles		x	p. 140
Lewis & Perry "Theoretical model of impact on instruction"	2014		LS cycle (Study, Plan, Do, Reflect)		intervening changes in teachers' knowledge, teachers' professional community norms and routines, and instructional materials, tools	improvement of instruction and student learning	I-P-O	Lewis et al. (2006); Desimone, (2009)	x	p.24

Norwich and Ylonen	2015	LS conditions and context	LS cycle (LS goals, pedagogy used, pupil characteristics, knowledge used, process complete)			learner outcomes, Teacher outcomes (quantitative), pedagogic outcomes (qualitative), Continue to use LS, School/ Department use of LS	I-P-O,		x	p. 635
Lewis "Lesson study cycle and pathways of impact: a theoretical model)	2016		LS cycle (study, plan, do, reflect)	study of materials and instruction with colleagues	pathways of impact: teachers' Knowledge, teachers' Beliefs and Dispositions, Teacher Learning Community Norms & Routines, Curriculum	instruction and student learning	I-P-O	Lewis et al. (2006)		p. 572
Norwich et al.	2016	subject knowledge, subject pedagogical knowledge, educational psychology, neuropsychology		reflective practice		enhanced teacher and learner outcomes	I-P-O	Ylonen & Norwich (2013); Norwich (2015)	x	p. 181
Bae et al. "Theoretical framework of teacher learning and change in lesson study"	2016	knowledge and beliefs, professional learning community, teaching and learning resources	LS cycle: planning, research lesson, reflection of student learning, investigation		knowledge and beliefs, professional learning community, teaching and learning resources		circular, iterative	Clarke and Hollingsworth (2002)		p. 168
Schipper et al.	2017	the change environment (school context), personal domain (knowledge, beliefs, and attitude)	LS domain (LS procedure, LS facilitator, professional experimentation)	enactment, reflection		domain of consequence (salient outcomes)	circular	Clarke and Hollingsworth (2002)	x	p. 301
Lewis et al. "Theoretical model of	2019		LS cycle (study, plan, teach, reflect)		intervening changes in teachers' knowledge, teachers' beliefs, routines and norms of	instruction and student learning	I-P-O	Lewis et al. (2006); knowledge integration		p. 15

lesson study impact"					professional learning, instructional tools and routines			environments, self-determination theory, and pedagogies of enactment		
Watanabe et al.	2019		CLR components, potential sources of learning (detailed description provided in model – some of the named components could be viewed as inputs or processes)		intervening changes and learning in TtP, CLR, mathematical knowledge, mathematical standards, pedagogical strategies and routines, proficiency in observing students, Students' mathematical thinking, etc.	two sequential outcomes: instructional improvement leading to improved student learning	I-P-O	Lewis et al. (2009)	x	p. 51
Akiba et al. "Conceptual model"	2019		duration (time span, amount), facilitator orientation (student thinking), material quality	effective inquiry process		positive changes in knowledge, self-efficacy, and expectations	I-P-O			p. 355
Mynott "Model of potential teacher learning outcomes in lesson study"	2019		LS work	cognitive dissonance		1. Absence of dissonance, 2. Dysfunctional dissonance, 3. Limited learning Dissonance, 4. Rich learning dissonance	reduced I-P-O structure	Mynott (2017)		p. 118
Dudley	2019	time, routines, habits, expectations		knowledge creation, knowledge application, knowledge systematization		knowledge creation, knowledge application, knowledge systematization	linear, not I-P-O		x	p. 214

Lee & Tan "The structure of the Lesson Study activity system"	2020	community (the school as PLC); rules (LS norms, PLT meeting norms), subject (the teachers in the PLT, instruments	division of labor (teachers participate in sessions, a teacher teaches a lesson, LS facilitators coordinate efforts, Level manager oversees administrative matters)	object (experimentation with teaching, understanding how pupils learn and culture of open classroom)		classroom outcomes (improved teaching and student learning)	mix between I-P-O and circular	based on Cultural-Historical Activity Theory	x	p. 10
da Ponte et al.	2022	external domain (source of information or stimulus)		domain of practice (professional experimentation), enactment and reflection	personal domain (knowledge, beliefs, and attitudes)	domain of consequence (salient outcomes)	circular	Clarke and Hollingsworth (2002)	x	p. 7, 10, 12

Note. If a publication referred to their model with a specific name, this name is noted in the column "study". For further information and context, the table provides the page number(s) on which the models can be found in their respective publications. Abbreviations used in table: CLR = collaborative lesson research, PLC = professional learning community, PLT = professional learning team, Ttp = teaching through problem solving.

The dimension of *process* refers to "interactions directed toward task accomplishment" (Mathieu et al., 2019, p. 18), meaning all actions that transform inputs into outputs. These actions are defined in 11 out of 18 models, however, they also differ between models and tend to remain abstract. The most frequently named process is reflection or related activities (i.e., inquiry process, review, dissonance). Four models refer to experimentation, enactment, or the application in the classroom (da Ponte et al., 2022; Dudley et al., 2019; Lee & Tan, 2020a; Schipper et al., 2017), and two to observation or an open classroom (Lee & Tan, 2020a; Ylonen & Norwich, 2013). One model includes the collaborative study of materials and instructions (Lewis, 2016). It has to be noted that sometimes processes are discussed within an article, yet they remain unspecified in the model itself. As the goal of models is to make underlying processes visible, this analysis regards the dimension of processes as unspecified in those cases.

Lastly, the dimension of output, similarly to the dimension of input, has been separated into two aspects: *area of impact* and *outcomes*. Output is defined as "results and by-products of team activity" and include the quality and quantity of participants' performance, as well as their reactions and satisfaction with their activities (Mathieu et al., 2019, p. 18). As Table 2 indicates, the majority of models define outcomes in relation to the affected areas. These areas include teachers' knowledge and their believes, their behavior, as well as their motivation and commitment. In addition, these models refer to outcomes such as the improvement of instructions, or improved instructions that lead to improved student learning. While some models illustrate this sequence of two outcomes, none of them include additional long-term outcomes that have been suggested by Guskey (2000), such as structural changes.

Concerning the overarching structures of the models summarized in Table 2, we can see that the majority of models (n = 14) follow an I-P-O structure. In fact, only three models are circular (Bae et al., 2016; da Ponte et al., 2022; Schipper et al., 2017), and one model appears linear but without a recognizable I-P-O design (Dudley et al., 2019). Considering the existing theories and prior models that have informed these LS models, it is striking that all three circular LS models are based on Clarke and Hollingsworth's (2002) conceptualization of teacher learning. Most models that have adopted an I-P-O design are, in contrast, based on the early model offered by Lewis et al.'s (2006). According to Lewis et al. (2006), their model is grounded in earlier efforts to explain teachers' instructional improvements through LS (Lewis et al., 2005). Their model suggests intervening changes in teachers' knowledge, their

commitment and motivation to improve, and their resources for teaching. While Lewis et al. (2005) base this approach on descriptions of Japanese LS written in the English language, it appears that their model is not based on a specific theoretical approach to teacher learning. Table 2 indicates that only two models that are based on the I-P-O structure are informed by theoretical approaches, namely cognitive dissonance (Mynott, 2019) and Cultural-Historical-Activity-Theory (Lee & Tan, 2020).

Lastly, Table 2 indicates that nine models have been developed or adapted for a particular piece of research. This means that half of all models are restricted in the sense that they include aspects that cannot easily be applied to a new context. The model by Watanabe et al. (2019), for example, defines outcomes that pertain specifically to their adaptation of LS and their use of LS within a specific project.

2.2.3. Limitations to Current Models

The review of models used to explain LS's mechanisms hints towards some limitations and implications for current research needs. As the systematic description offered in Table 2 demonstrates, the field of LS has used and produced several models in the past two decades. A shared model or perspective appears, however, to be currently missing from the literature. None of the reviewed models seem to have been conceptualized with the aim to provide such a shared perspective. The continued development of models that inform a specific piece of research, however, suggests that a common conceptualization that can be transferred across contexts and inform LS research regardless of its specific focus, is much needed.

Specifically, the analysis showed that there is no coherent understanding in the field of LS of any of the dimensions of input, process, or output. Only few models described input factors beyond the surface features of LS, and the conceptualization of processes differed widely. The dimension of output, while consistently present in models, usually only recognized teachers' instructions as the desired PD outcome, with some models also including students' improvements. This means that current models cannot account for immediate, mid-term, or long-term outputs, despite LS being classified as an iterative and long-term PD approach (Lewis, 2015; Yurkofsky et al., 2020). This limitation is exaggerated by the structure of most models, which follows the linear I-P-O design and does not display iteration.

This lack of a shared conceptualization of LS appears to be the root of several problems discussed in connection with the first research need, the knowledge base on LS. The

systematic review by Seleznyov (2019), for instance, found that only a minority of studies have investigated LS outcomes that expend beyond teachers' immediate reactions. Both Cheung and Wong (2014) and Willems and van den Bossche (2019) noted in their reviews that LS research tends to employ incoherent outcome measures for the success of LS. In light of the present analysis, these findings are not surprising, given that a shared model that specifies a range of outcomes is currently missing.

The research need identified by Lewis et al. (2006) to explain LS's underlying mechanisms has therefore only been advanced in parts. Reasons for this limited progress might be the incoherent use of theories to inform these models. We saw, for instance, that only a small minority of models explicitly grounded their approach in previous models or theories, such as Clarke and Hollingsworth's (2002) interconnected model of teacher learning. While the three LS models building on Clarke and Hollingsworth were able to account for inputs and processes, the dimension of output did not offer any conceptualization of tangible outcomes over time. In contrast, those models based on Lewis et al.'s (2006) approach usually defined specific areas of impact as well as sequential outcomes of improved instructions that lead to students' achievements. Inputs, however, were mostly constricted to LS features, and processes remained largely undefined. This can be explained by the lack of theoretical backdrop that could provide answers regarding, for example, the processes that underpin and connect these dimensions. The lack of a shared conceptualization of LS and its outcomes might also explain why several reviews (see Table 2, Appendix) found that studies on the effectiveness of LS use inconsistent outcome measures.

The analysis therefore indicates that the use of theories, or lack thereof, shape the structure of models. All three models that build on Clarke and Hollingsworth (2002) have adopted a circular design, while other models follow a linear structure and necessarily fail to account for iteration. This observation connects to Boylan et al.'s (2018) argument that models and theories are developed with a specific goal in mind and cannot account for everything. General theories of learning therefore struggle to classify tangible or local outcomes, while specific models cannot be widely applied.

Given that a shared LS model is currently missing, Lewis et al.'s (2006) second research need should remain a priority in LS research. This analysis re-emphasizes the critical importance of a shared theory-informed LS model that can provide conceptual coherence to the field. It further raises several open questions that can promote the development of such

a model: (1) Which criteria would a model of teacher learning through LS need to fulfil, so that it can be used across contexts?, (2) How can we develop a model that is concrete but still widely applicable?, (3) How can such a model address all three dimensions of input, process, and output, and also specify iteration?, (4) How would such a model inform both research and practice?

2.3. Research Need 3: Design-Based Research Cycles

The third critical research need identified by Lewis et al. (2006) was to improve our understanding as well as our implementations of LS by means of design-based research cycles. The aim of this section is to first define relevant terminology, such as design-based research (DBR), action research (AR), educational research, classroom-research, and improvement science. Then, the use of DBR and AR in LS research, as well as the differences between these methodologies and LS, are discussed. Lastly, this section raises some open questions and implications regarding this research need.

2.3.1. Practice-Based Approaches to Study Professional Development

The golden standard to investigate a PD's effectiveness is traditionally considered to be a controlled and experimental design (Bryk, 2015). Given that the improvement of schools on a national level requires substantial financial resource, those responsible for providing these funds seek to base their decisions on the best possible evidence for effective policies (Pogrow, 2017). In the US, this led to the establishment of *What Works Clearinghouse* (i.e., WWC; Bryk, 2015; Kratochwill et al., 2013; Pogrow, 2017) around the turn of the century. The WWC provides standards for the assessment of educational innovations and officially validates those that succeed. In recent years, however, there has been a shift away from the *What Works*-paradigm, as the positive impact that evidence-based policies have in a certain context is often challenging to replicate in another (Bryk, 2015; Pogrow, 2017). Alternative practice-based methodologies that support the transfer of innovations from research into the classroom have since gained popularity, most notably DBR and AR (Anderson & Shattuck, 2012; Bryk, 2015; Tinoca et al., 2022).

DBR can be understood as a research methodology "designed by and for educators" (Anderson & Shattuck, 2012, p. 16) that relies on iterative cycles of implementation as a way to understand local processes and react to emerging problems. These improvement cycles usually follow the Plan-Do-Study-Act structure (i.e., PDSA; Cohen-Vogel et al., 2015). In other words, a specific intervention is planned and then tested and assessed in the field. The intervention is then either abandoned or modified and enhanced based on the data from this implementation. If revised, the intervention is tested again in the field (Cohen-Vogel et al., 2015). The focus of DBR thereby lies not only on the effectiveness of the educational

intervention itself, but on its practical design principals; meaning those conditions and features necessary for an intervention to work in diverse settings (Anderson & Shattuck, 2012). Importantly, DBR is not a research method in itself, but an approach to research that makes use of both quantitative and qualitative research methods, such as pre- and post-tests, interviews, field notes, or video and audio taping, such as pre- and post-tests, interviews, field notes, or video and audio taping (Tinoca et al., 2022).

The term DBR is sometimes used as an umbrella term in the literature, or used synonymously with other practice-based concepts, such as AR (Nijhawan, 2017). Similar to DBR, AR is also based on iterative PDSA cycles and is conducted by educators in their classrooms (Willis & Edwards, 2014). In their review, Anderson and Shattuck (2012) argue, however, that the two approaches can be clearly separated from each other according to the roles that educators take on. In DBR, educators provide practical insights from the classroom, but do not actively conduct research. Researchers and educators nevertheless share the responsibility for their work. In AR, educators usually take on both roles: they act as teachers and researchers. This means that in AR the partnership between researchers and educators, as well as the researchers' expertise, is usually lacking (Anderson & Shattuck, 2012). The two approaches also differ in their objectives: The goal of DBR is to both impact practice and inform theories, while the goal of AR is usually restricted to the impact on practice (Anderson & Shattuck, 2012).

These diverging objectives are mirrored in the ways that DBR and AR are conducted. DBR usually follows the procedures of educational research, while AR is best described as classroom-research. Nunan (2005), addressing this difference, explains that the term educational research refers to the "application of the scientific method to educational topics, phenomena, or questions in search of answers" (p. 6). By scientific method, he refers to a methodical, highly-structured, and step-by-step process. Classroom-based action research, on the other hand, is defined as an effort to investigate and improve a specific classroom problem at a contained site, such as a specific school (Nunan, 2005). The aim of this process is for teachers to examine their own practice and to develop their own professional competence by means of critical reflection. Importantly, AR provides teachers with a defined structure that can facilitates this inquiry process (Nunan, 2005).

This difference is also illustrated in Elliott's (1991) influential definition of AR, which posits that knowledge generation in AR "depends not so much on 'scientific' tests or truth" but

rather on the "usefulness in helping people act more intelligently and skillfully" (p.69). The aim of AR is therefore not to test and validate scientific theories. Instead, teachers develop practice theories and test them in the classroom (Feldmann, 2018). DBR and AR thus share certain similarities, but their application, participants, and objectives differ.

The establishment of both DBR and AR has gone hand in hand with the aforementioned shift in paradigm that is still gaining traction in the field of education: the shift away from the *What Works*-paradigm and towards improvement science (Bryk, 2015; Cohen-Vogel et al., 2015). Improvement science—also referred to as *continuous improvement* (Cohen-Vogel et al., 2015)—has been defined in various ways in the literature, but is usually regarded as an approach or a field of study, rather than research in itself (Park et al., 2013). Park et al. (2013) describe improvement science as an effort to identify those methods, approaches, and theories that facilitate the improvement of a specific innovation or intervention within a specific context. The PDSA cycle is an important tool in improvement science (Cohen-Vogel et al., 2015) as way to generate "practice-based evidence" (Bryk, 2015, p. 469).

While scholars frequently dispute over the precise definitions and classifications of the discussed approaches, this review of the literature suggests that improvement science, as an area of study, includes the methodologies of DBR and AR. Both DBR and AR reflect separate approaches to conducting improvement science, and each approach adheres to distinct standards and necessitates the use of certain quantitative and qualitative techniques. The next subsection addresses the similarities and differences between improvement science and LS and explores empirical research studies that have employed improvement cycles as a way to study LS.

2.3.2. Improvement Science and Lesson Study

Improvement science and methodologies such as DBR and AR have been addressed in the field of LS several times (e.g., Dick et al., 2022; Elliott, 2019; Lewis, 2015;). As Hanfstingl et al. (2019) note, however, there is a "striking lack of clearness" (p. 456) concerning these concepts in the LS literature. This lack of clearness seems to stem from the fact that LS is sometimes treated as the object of research (i.e., LS is an intervention that is studied), and sometimes as a form of classroom-research itself (i.e., LS is the methodology through which teachers investigate a specific area of instructions).

Lewis (2015), for instance, describes LS as an example of improvement science. Specifically, she argues that LS in Japan is an illustration of how educational innovations can be scaled-up through improvement science on a grand scale. LS in Japan is conducted on multiple layers of the educational system, including schools, districts, universities, and professional associations (Lewis, 2015). The findings of LS cycles are typically disseminated within and across all these layers. This way, instructional innovations can inform curriculum development and spread widely (Lewis, 2015). It should be highlighted that Lewis' (2015) example refers to the dissemination of instructional innovations through teachers' ongoing involvement in LS, rather than the spread and refinement of the LS approach itself. LS is thus described as a tool or methodology of improvement science.

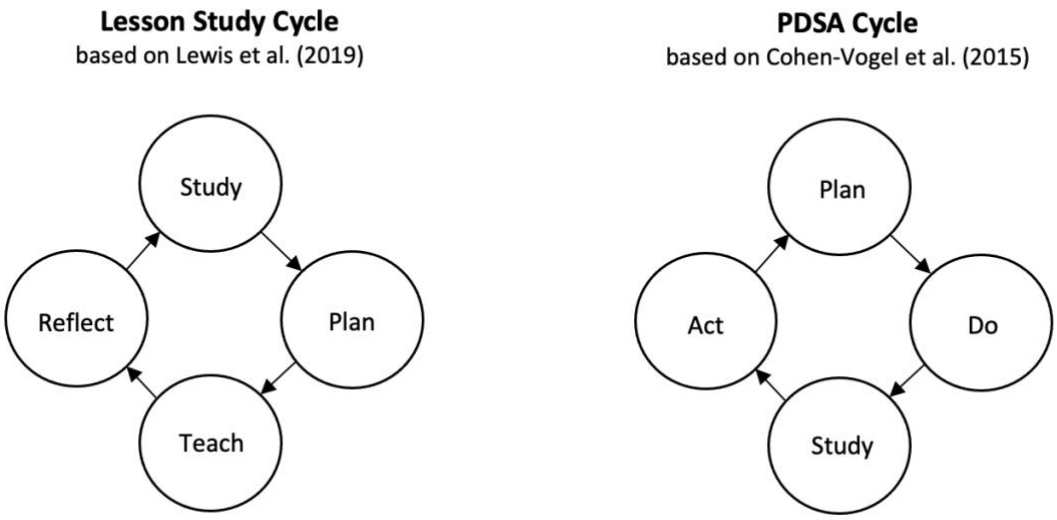
Regarding DBR, the literature provides some examples of empirical studies that have used DBR in order to better understand how LS can work in a specific context (i.e., LS is the research object that is investigated through the use of an improvement methodology). These examples include, for instance, the studies by Brown et al. (2016) and Norwich & Ylonen (2013, 2015). Groves et al. (2016) also categorize their research approach as DBR, but seem to conceptualize the LS implementation itself as part of this research approach. Their findings relate to both the effectiveness of LS as a PD model in the Australian context and to a specific area of instruction that teachers investigated in their LS cycles. The literature offers additional studies that could be identified as using a form DBR, but categorize their own approach in different terms, such as *developmental evaluation* or *participatory research* (e.g., Godfrey et al., 2019). The study by Dudley et al. (2020), on the other hand, reports on the use of AR to establish LS in a London district, UK. The authors explain that the strong partnership between educators and researchers played a critical role in the process. Their understanding of AR thus seems to disagree with the one of Anderson and Shattuck (2012), who posit that researchers are largely absent in AR. According to the definitions discussed in the previous subsection, Dudley et al.'s (2020) approach would probably be categorized as DBR instead.

Turning to AR in the field of LS, the use of terminology becomes even less clear. Baumfield et al. (2022), who conducted a systematic review into the use of LS in initial teacher education, note that the majority of eligible studies categorized themselves as AR. This is in line with Yoshida et al.'s (2021) assessment of AR as "one of the most widely accepted research methodologies for theoretical (and even practical) improvement of Lesson Studies" (p. 33). Their description of AR, however, clashes with the prevalent definition of AR as a way to first

and foremost impact practice, not theory (e.g., Elliott, 1991; Nunan, 2005). It therefore seems that multiple understandings of AR exist in the field of LS. In fact, AR has also been equated with LS (e.g., Dudley, 2014; Pérez Granados et al., 2022). Dudley (2014), for instance, makes the case that Japanese LS can be considered an early type of what is now known as AR in the West. When we compare the PDSA cycle with the LS cycle side by side (see Figure 3), this comparison is quite clear. Although the order differs, the activities in AR and LS are similar to each other: teachers develop a specific objective, they test ideas in the classroom, they study the curriculum, and they base their decisions on the analysis of their observations or other form of data.

Figure 3

The Lesson Study Cycle and the PDSA Cycle Side by Side



In an effort to identify differences between LS and AR, Stylianou and Zembylas (2019) remark that, unlike LS, AR is not confined to the improvement of instructions and does not necessarily rely on outside expertise or a live research lesson. They add that the use of a methodology such as AR and a range of methods to collect different kind of data provides a kind of legitimacy that LS cannot offer. This comment by Stylianou and Zembylas (2019) raises several key questions, such as (1) to what extend are the conclusions teachers draw in LS valid and the product of evidence-based reasoning?, and (2) should LS be conceptaulized as a form of research (i.e., teachers need to adhere to specific standards in their inquiry process) or

rather a form of PD (i.e., teachers should first and foremost develop their own professional competence)?

The latter question has been problematized by Takahashi and McDougall (2016) in the US context. They introduced an adapted form of LS, *Collaborative Lesson Research* (CLR), that emphasizes teachers as researchers. Specifically, teachers define a clear "research purpose" and the goal of the research lesson is not to enact the lesson plan, but to "search for a solution to a teaching-learning problem" (Takahashi & McDougall, p. 519). During the study stage, teachers write a so-called research proposal, in which they detail their rational behind their research purpose and explain the way in which they aim to investigate this purpose. CLR is typically supported by one or two knowledgeable others—such as researchers—, who may provide feedback to the proposal and support the LS team in each stage. The goal of the post-lesson discussion is to analyze the data collected during the research lesson in respect to the research purpose. Lastly, the findings are shared with a larger community, for example by inviting people to the live research lesson or publishing the research proposal and the team's reflections. This form of LS therefore highlights the research process and formulates stricter structures for how teachers conduct LS. Takahashi and McDougal's (2016) adaptation of LS, and the re-branding as Collaborative Lesson Research, can therefore be seen as an effort to take a clearer position concerning their understanding of LS.

This section has highlighted the inconsistent terminology in the field of LS concerning DBR, AR, and LS. The inconsistent ways in which publications categorize their own research approach suggest that the issue does not lie with terminology per se, but rather with conceptual differences concerning the scope of LS. The next section addresses the ways in which conceptual ambiguity hinders progress in the field of LS.

2.3.3. Limitations to the Current Use of Improvement Science

The review of the literature indicates that, as suggested by Lewis et al.'s (2006), design-based improvement cycles have been employed in the field of LS in the form of both DBR and AR. Several scholars have remarked, however, that a shared understanding of what distinguishes DBR, AR, and LS is lacking in the field (e.g., Hanfstingl et al., 2019; Elliott, 2019). We saw, for instance, that LS is sometimes regarded of as a PD and other times as a methodology akin to AR. This ambiguity surrounding the scope of LS raises several issues and questions.

Lewis et al.'s (2006) recommendation to employ design-based cycles to study LS is remains valid today. The idea of improvement science offers two distinct affordances to the study of effective PD. The continuous improvement of an educational intervention generates valuable insights into the key activities and protocols that make the intervention effective for a certain setting (Bryk, 2015). In addition, improvement science has the ability to promote the collaboration between researchers and educators and ensure that theory-based innovations are successfully transferred into classrooms. As Guskey (2017) remarks, teachers typically find results that they can observe for themselves in the classroom more convincing that results published in research studies. DBR and AR have the potential to accommodate this need to actually experience how innovative strategies affect student learning. DBR in particular has the potential to bridge the gap between research and practice, as it generally conforms to stricter criteria than AR (Anderson & Shattuck, 2012) and represents a research methodology to investigate LS as the research object.

This dissertation found, however, that there is room for improvement concerning how DBR and AR are currently employed in the field of LS. To date, the issue of how we investigate LS, albeit discussed by some scholars, has not yet gained significant attention in the field. Several systematic reviews have identified problems in the ways LS publications employ and report research methodologies (e.g., Seleznyov, 2019; Willems & van den Bossche, 2019). Yet, a critical discussion that puts this issue in the spotlight is currently missing.

This discussion could advance the field in several ways. First, a discourse on how we investigate LS could help establish a shared understanding of the terms DBR, AR, and LS, as well as raise awareness for conceptual differences in the field. Clear terminology may support the field in capitalizing on improvement methodologies. In addition, it could spark a debate on the quality of research in the field LS and bring attention to the fact that publications should clearly position themselves as either educational research or classroom-based research, so that the research community knows how to understand the findings and engage with them. Finally, a systematic review of studies based on DBR could shed some light on, for instance, the kind of information DBR studies contribute to the question of LS's effectiveness, or the extent to which findings from DBR can be compared and synthesized. This information would be of great value to the field, because much like controlled research-designs, DBR is subject

to several limitations. DBR requires lengthy and potentially expensive studies and is necessarily more susceptible to researchers' biases than other more controlled designs (Anderson & Shattuck, 2012). In addition, both Tinoca et al. (2022) and Limere et al. (2017) remark that for improvement cycles to be recognized as a trustworthy and replicable research approach, they must be meticulously recorded in publications.

To sum up, Lewis et al.'s (2006) research need to improve our understanding of LS through design-based research cycles to has only been partially addressed in the field. There are several questions that could promote progress on this issue: (1) How are DBR and AR currently employed in LS research?, (2) Which design-principals for LS have been identified in DBR publications, and how do they compare to each other?, (3) How are DBR and AR currently documented in research publications?, and (4) Which tools could be used (and re-used) to document DBR and AR in a comparable way?

3. Synthesizing Current Research Needs

The past 17 years have witnessed a huge growth in research on LS, and much progress has been made. Figure 2 illustrated that an increasing number of literature reviews and systematic reviews of LS research has been conducted. This development can be seen as a sign that the field of LS is maturing, given that the goal of reviews of the literature is to offer "a new vantage point" (Alexander, 2020, p. 6) to the field. The first part of this dissertation, seeking to offer such a new vantage point, presented a synthesis of existing reviews on LS in order to take stock of the current research needs in the field. Despite the developments in the field, this synthesis argues that the research needs identified by Lewis et al. in 2006 have only been advanced in part and need to remain a priority in the field. Importantly, the three research needs were formulated with reference to the arrival of LS to the US. Given that LS has since spread far beyond the US context, the present analysis implies that each research need should to be updated and revived (see Table 3).

Table 3

Overview of Critical Research Needs in LS: Then and Now

	Research Need 1	Research Need 2	Research Need 3
Then Research needs in LS according to Lewis et al. (2006)	**Expansion of the Descriptive Knowledge Base of Japanese and U.S. Lesson Study** to expand the descriptive knowledge base on LS in an effort to describe LS's characteristics and determine adaptations pertinent to LS's implementation in US settings	**Explication of the Innovation Mechanism** to investigate the mechanisms through which teachers learn in LS and develop a model that represents these mechanisms as well as LS surface features	**Design-Based Research Cycles** to use design-based research cycles to improve LS adaptations and support theory-building
Now Research needs in LS derived from the current literature	**Development of an international, coherent, and rigorous knowledge base on Lesson Study** to build an international research base that - uses theory-based and rigorous approaches to study teacher learning - provides transparent descriptions of their research methodologies and LS intervention	**Development of a conceptual model of Lesson Study** to develop a theory-based conceptual model of LS that - can be used across contexts - systematically describes inputs, processes, and outputs - can be applied widely	**Development of a understanding of DBR** to capitalize on improvement methodologies by - establishing clear terminology - synthesizing current use and findings of DBR in LS - developing and testing protocols that can guide DBR

Research Need 1. The first research need should be updated to the development of an international, coherent, and rigorous knowledge base on LS. The synthesis of literature and systematic reviews on LS reveals that the rapidly growing knowledge base on LS suffers from several shortfalls. First, the issue of who contributes knowledge to the current research base has not been investigated in depth and it remains unclear to what extend and in which roles educators participate in the generation of knowledge. Second, the ways in which knowledge is being produced and systematized can be improved. Crucial information on research methodology or the LS intervention are persistently missing or insufficiently reported in publications. The lack of shared standards concerning how to best report LS research prevents both the synthesis and replication of research findings. The field should therefore focus on generating a knowledge base that adheres to clear standards that enable the synthesis, verification, and replication of both quantitative and qualitative research. These standards need to be derived from the wider discourse on quality in science and educational research and be adjusted to the field of LS.

Research Need 2. The second research need should be updated to the development of a conceptual model of Lesson Study. The scoping of the literature demonstrates that a number of models have been developed or adapted to explain teacher learning in LS. A critical comparison of these models with each other as well as with general models of teacher learning and classifications of learning outcomes, however, reveals some drawbacks that need to be addressed. First, current models are predominantly tailored towards individual pieces of research or cannot be easily transferred across settings due to their restricted scope. Second, the majority of current models is fragmented, meaning that they neglect some of the dimensions necessary to provide a complete picture of learning through LS (i.e., inputs, processes, and outputs). Third, the analysis indicates that some models lack a clear theoretical footing, while other models are limited in their scope due to their theoretical underpinning. To sum up, the field of LS still lacks a model that can be used across contexts, systematically describes inputs, processes and outputs, and is informed by contemporary perspectives on teacher learning and PD.

Research Need 3. The third research need should be updated to the development of a shared understanding of DBR**.** The review of the literature indicates that design-based research cycles offer several benefits to the study of LS implementations, yet the full potential of such methodologies has not yet been unlocked. First, the usefulness of improvement

methodologies has to date been discussed mainly on the sidelines. Second, despite the fact that the empirical LS literature offers several examples of DBR and other improvement methodologies, the distinctions between these methodologies are hazy and a shared conceptualization of DBR, AR, and LS is missing. In order to capitalize on the ideas of improvement science, the field should strive towards a clear use of terminology, a synthesize of how DBR is currently used in research studies and what we can learn from the findings of these studies, as well as the establishment of tools or protocols that can guide DBR and AR on LS and improve the quality and comparableness of small-scale and contextualized LS research.

4. Deriving Research Questions for this Dissertation

This dissertation argues that progress in the field of LS and the advancement of the critical research needs in the field are stifled by several limitations. Based on the review of the literature, these limitations are (1) the frequent lack of comparable and replicable descriptions of the LS intervention in publications, (2) the incoherent use or lack of use of theoretical frameworks to explain teacher learning through LS, (3) the inconsistent use of terminology and concepts, and (4) the lack of scientific rigor in research studies and of established ways or tools to measure the effectiveness of LS (Larssen et al., 2018; Seleznyov, 2019; Willems & van den Bossche, 2019; Xu & Pedder, 2014). These limitations do not affect an individual research need, but they affect all three and cannot be considered separately. For instance, the lack of a theory-based conceptual model of LS (Research Need 2) and the lack of a shared understanding of DBR and AR (Research need 3) both influence the quality of the knowledge base on LS (Research Need 1).

This dissertation aims to make progress on the current research needs by examining the extent and nature of these limitations in three research papers. As the thorough examination of each limitation is, however, beyond its scope, this dissertation focuses on limitations (1), (2), and (3). The findings are nevertheless expected to generate insights into all four limitations and jointly inform implications for future research needs. The starting point of the three research papers is the quest to develop a conceptual model for LS that can guide future research on LS. An important stepping stone in this effort is the theorization of the processes that teachers engage in when conducting LS. Given that teachers' observations and reflections have been identified as crucical in this process, yet at the same time they tend to be insufficently theorized and described in the literature (Larssen et al., 2018; Saito, 2012; Xu & Pedder, 2014), this dissertation primarily examines these two stages.

The first study provides an empirical and mixed-method analysis of how LS teams reflect together. Building on this account, the second article offers a systematic review of how the observation and reflection stages are theorized and reported in LS publications. The third article, taking a meta-perspective, proposes a shared conceptual model for the field of LS. The specific research questions and methodological approaches of each study are the following:

Study 1, *"We were thinking too much like adults": Examining the development of teachers' critical and collaborative reflection in lesson study discussions*: This mixed-method article discusses and reviews the theoretical concept of critical and collaborative reflection by re-

framing the reflection stage of LS in Korthagen and Vasalos' (1985, 2005) reflection model. In order to examine how critical and collaborative reflection is enacted by LS teams, audio-recordings of post-lesson discussions were collected at four German primary schools. The audio-recordings were a transcribed and coded using a coding tool developed based on the ALACT model. Concerning the quantitative analysis, the frequencies and distributions of codes over the time-span of each discussion were compared between schools and between code categories. Concerning the qualitative analysis, excerpts from schools were used to illustrate the different ways in which LS teams navigated their reflections, responded to difficulties, and formulated solutions. The specific research questions addressed in this study were: (1) How do LS teams differ in the depth of their reflection in terms of reflective stages?, and (2) how do LS teams' respective trajectories through their reflective practice differ on a micro-diachronic scale?

Study 2, *A Systematic Review of Transparency in Lesson Study Research: How Do We Report on the Observation and Reflection Stages?*: Previous research has established that publications on LS frequently lack key information that is necessary to fully comprehend a study or replicate its findings. This review draws on Moravcsik's (2020) three dimensions of research transparency in order to establish which information concerning the observation and reflection stages needs to be reported in LS research. Using a newly developed and pre-registered coding protocol (Kager et al., 2021), 129 research articles on LS published between 2015 and 2020 were reviewed, assessed, and coded. Based on the findings, a checklist was developed. This checklist may act as a framework for future LS articles to increase the transparency of their LS intervention. This systematic review aimed to answer the following questions: (1) How transparent are LS articles in reporting their observation and reflection stages?, and (2) which theoretical frameworks are used in these studies to conceptualize the observation and reflection stages in LS?

Study 3, *A conceptual model for teachers' continuous professional development through Lesson Study: Capturing inputs, processes, and outcomes*: This conceptual article aims to develop a conceptual model for the field of LS that connects diverse LS implementations and their findings to a common schematic framework that provides coherent terminology and explanatory power. The specific scope and requirements of such a model were determined by reviewing the research base on teacher learning, PDs, and organizational psychology. As a next step, existing LS models were analyzed and gaps in their utility identified. These gaps were

bridged by proposing a new and extended LS model that combines concrete and theory-led inputs, processes, and outputs. The article then illustrates how this model can be applied by both researchers and educators.

5. Context of this Dissertation

This dissertation was conducted within the "Leistung macht Schule (LemaS)"-initiative (["Excellence in School Education"], BMBF & KMK, 2016). The initiative, organized by the German Research Association of the Federal Ministry of Education, is a nationwide project running from 2018 to 2023 that includes 22 component projects working with overall 300 schools (Weigand, 2022). The main goal of the initiative is to create innovative concepts that support high-achieving and potentially high-achieving students within the regular classroom. In our component project, we introduced LS to 19 primary schools across Germany. The objective of our component project is twofold: to develop teachers' instructions and knowledge in regard to high-achieving and potentially high achieving students through their participation in LS, and to investigate how LS can be an effective approach to PD at German schools (Jurczok et al., 2020).

LS is relatively unknown within the German education system and the public, yet interest in LS as a PD seems to be increasing. LS has been transferred to various school settings in Germany several times in connection with projects or research initiatives (e.g., Knoblauch, 2017). Some scholars have also discussed LS as a potentially beneficial approach to PD in Germany (e.g., Gervé, 2007; Kager et al., 2022; Klopsch & Sliwka, 2021; Kullmann, 2012). However, to date only few research articles on LS in Germany have been published in English and are accessible to the international research community (e.g., Hallitzky et al., 2021; Yoshida et al., 2021). To the best of our knowledge, the LS-project within the LemaS-initiative represents the largest effort to transfer LS to German schools and examine its usefulness and effectiveness.

As Richter and Richter (2020) remark, teachers in Germany gain the essential professional competences during teacher training at university, but are legally required to continue their professionalization throughout their careers. While a range of PD offers exist, the majority of teachers participate in only one to two PD offers per year, which are usually unconnected and do not last longer than a day (Richter & Richter, 2020). It therefore seems that a PD approach such as LS could be a way to engage teachers in Germany in continuous PD over a longer period of time. In addition, Jurczok et al. (2020) suggest that the transfer of LS to German schools could provide a viable framework for teachers to intensify their collaboration. Massenkeil and Rothland (2016) showed that German teachers seldomly engage in collaboration that goes beyond the sharing of materials. This is at odds with findings from a

study by Richter and Pant (2016), which indicate that the vast majority of teachers in Germany regard collaboration as important. The necessary preconditions and structures to engage in complex forms of collaboration, such as the joint development of lesson plans or conducting open lessons, are, however, not always available (Richter & Pant, 2016). Jakobeit et al. (2021) examined whether schools in Germany can offer the preconditions and resources necessary to implement LS, such as sufficient staff, materials, space, time, and support by the school leaders and from external experts. The analysis of these preconditions at 135 schools that participate in the LemaS-project suggests that most schools lacked at least some of these resources, yet a cluster of 22 schools reportedly managed to establish structures for collaboration despite this lack (Jakobeit et al., 2021).

Together, these findings suggest that there is a need for PD approaches in Germany that can offer a platform for teachers to intensify their collaboration and learn together over a sustained period of time. The international research base indicates that LS, if adapted to the requirements of the German educational system, could offer such a platform. This dissertation, written in the context of the LemaS-project, contributes to research on the transfer of LS to Germany in two ways. The first research study of this dissertation provides a detailed documentation and analysis of how teachers at German schools engage in collaborative and critical reflection. In addition, this dissertation aims to derive considerations for LS research in Germany based on the assessment of the international research base.

6. References

Akiba, M., Murata, A., Howard, C. C., & Wilkinson, B. (2019). Lesson study design features for supporting collaborative teacher learning. *Teaching and Teacher Education*, *77*, 352–365. https://doi.org/10.1016/j.tate.2018.10.012

Albers, B., & Pattuwage, L. (2017). *Implementation in education—Findings from a scoping review*. Centre for Evidence and Implementation - Prepared for Evidence for Learning (E4L). https://doi.org/10.13140/RG.2.2.29187.40483

Akiba, M., Murata, A., Howard, C. C., & Wilkinson, B. (2019). Lesson study design features for supporting collaborative teacher learning. *Teaching and Teacher Education*, *77*, 352–365. https://doi.org/10.1016/j.tate.2018.10.012

Anderson, T., & Shattuck, J. (2012). Design-based research: A decade of progress in education research? *Educational Researcher*, *41*(1), 16–25. https://doi.org/10.3102/0013189X11428813

Bae, L. C., Hayes, K. N., Seitz, J., O'Connor, D., & DiStefano, R. (2016). A coding tool for examining the substance of teacher professional learning and change with example cases from middle school science lesson study. *Teaching and Teacher Education*, *60*(2016), 164–178. https://doi.org/10.1016/j.tate.2016.08.016

Bakkenes, I., Vermunt, J. D., & Wubbels, T. (2010). Teacher learning in the context of educational innovation: Learning activities and learning outcomes of experienced teachers. *Learning and Instruction*, *20*(6), 533–548. https://doi.org/10.1016/j.learninstruc.2009.09.001

Bandura, A. (1969). Social-learning theory of identificatory processes. In D. Goslin (Ed.), *Handbook of socialization theory and research* (pp. 213–262). Rand McNally & Company. https://tiny.link.bandura

Baumfield, V., Bethel, A., Boyle, C., Katene, W., Knowler, H., Koutsouris, G., & Norwich, B. (2022). How lesson study is used in initial teacher education: An international review of literature. *Teacher Development, 26*(3), 356-372. https://doi.org/10.1080/13664530.2022.2063937

BMBF, & KMK. (2016). *Gemeinsame Initiative von Bund und Ländern zur Förderung leistungsstarker und potenziell besonders leistungsfähiger Schülerinnen und Schüler. [Joint initiative by the federation and federal states for the support of gifted and potentially gifted pupils].* https://www.bmbf.de/de/leistung-macht-schule-3641.html

Borko, H., Jacobs, J., & Koellner, K. (2010). Contemporary approaches to teacher professional development. In P. Peterson, E. Baker & B. McGaw (Eds.), *International Encyclopedia of Education* (Vol.7, pp. 548-556). Elsevier.

Boylan, M., Coldwell, M., Maxwell, B., & Jordan, J. (2018). Rethinking models of professional learning as tools: A conceptual analysis to inform research and practice. *Professional Development in Education*, *44*(1), 120–139. https://doi.org/10.1080/19415257.2017.1306789

Brosnan, A. (2014). Introducing lesson study in promoting a new mathematics curriculum in Irish post-primary schools. *International Journal for Lesson and Learning Studies*, *3*(3), 236–251. https://doi.org/10.1108/IJLLS-09-2013-0050

Brown, C., Taylor, C., & Ponambalum, L. (2016). Using design-based research to improve the lesson study approach to professional development in Camden (London). *London Review of Education*, *14*(2), 4–24. https://doi.org/10.18546/LRE.14.2.02

6. References

Bruce, C. D., Flynn, T. C., & Bennett, S. (2016). A focus on exploratory tasks in lesson study: The Canadian "math for young children" project. *ZDM: The International Journal on Mathematics Education*, *48*(4), 541–554. https://doi.org/10.1007/s11858-015-0747-7

Bryk, A. S. (2015). 2014 AERA Distinguished lecture: Accelerating how we learn to improve. *Educational Researcher*, *44*(9), 467–477. https://doi.org/10.3102/0013189X15621543

Burkhardt, H., & Schoenfeld, A. H. (2003). Improving educational research: Toward a more useful, more influential, and better-funded enterprise. *Educational Researcher*, *32*(9), 3–14. https://doi.org/10.3102/0013189X032009003

Burrows, S. (2022). What are the key principles of lesson study?: A systematic review of the evidence. *International Journal of Teacher Education and Professional Development 5*(1), 1–14. https://doi.org/10.4018/IJTEPD.295548

Cajkler, W., Wood, P., Norton, J., & Pedder, D. (2014). Lesson study as a vehicle for collaborative teacher learning in a secondary school. *Professional Development in Education*, *40*(4), 511–529. https://doi.org/10.1080/19415257.2013.866975

Callahan, C. (2019). Middle school geography teachers' professional development centered around historical photographs. *Journal of Social Studies Research*, *43*(4), 375–388. https://doi.org/10.1016/j.jssr.2018.11.003

Cammarata, L., & Haley, C. (2018). Integrated content, language, and literacy instruction in a Canadian French immersion context: A professional development journey. *International Journal of Bilingual Education and Bilingualism*, *21*(3), 332–348. https://doi.org/10.1080/13670050.2017.1386617

Canonigo, A. M. (2016). Using a non-coercive process to engage mathematics teachers in lesson study. *International Journal for Lesson and Learning Studies*, *5*(4), 329–347. https://doi.org/10.1108/IJLLS-02-2016-0004

Century, J., & Cassata, A. (2016). Implementation research: Finding common ground on what, how, why, where, and who. *Review of Research in Education*, *40*(1), 169–215. https://doi.org/10.3102/0091732X16665332

Chen, X., & Zhang, Y. (2019). Typical practices of lesson study in East Asia. *European Journal of Education*, *54*(2), 189–201. https://doi.org/10.1111/ejed.12334

Cheung, W. M., & 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*, *3*(2), 137–149. https://doi.org/10.1108/IJLLS-05-2013-0024

Chi, M. T. H. (2008). Three types of conceptual change: Belief revision, mental model transformation, and categorical shift. In S. Vosniadou (Eds.), *International handbook of research on conceptual change* (pp. 61–82). Routledge. https://doi.org/10.4324/9780203154472

Chikamori, K., Ono, Y., & Rogan, J. (2013). A lesson study approach to improving a biology lesson. *African Journal of Research in Mathematics, Science and Technology Education*, *17*(1–2), 14–25. https://doi.org/10.1080/10288457.2013.826967

Chokshi, S., & Fernandez, C. (2004). Challenges to importing Japanese lesson study: Concerns, misconceptions, and nuances. *Phi Delta Kappan*, *85*(7), 520–525. https://doi.org/10.1177/003172170408500710

Chong, W. H., & Kong, C. A. (2012). Teacher collaborative learning and teacher self-efficacy: The case of lesson study. *The Journal of Experimental Education*, *80*(3), 263–283. https://doi.org/10.1080/00220973.2011.596854

Clarke, D., & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education*, *18*(8), 947–967. https://doi.org/10.1016/S0742-051X(02)00053-7

Coenders, F., & Verhoef, N. (2019). Lesson study: Professional development (PD) for beginning and experienced teachers. *Professional Development in Education*, *45*(2), 217–230. https://doi.org/10.1080/19415257.2018.1430050

Cohen-Vogel, L., Tichnor-Wagner, A., Allen, D., Harrison, C., Kainz, K., Socol, A. R., & Wang, Q. (2015). Implementing educational innovations at scale: Transforming researchers into continuous improvement scientists. *Educational Policy*, *29*(1), 257–277. https://doi.org/10.1177/0895904814560886

Coldwell, M. (2017). Exploring the influence of professional development on teacher careers: A path model approach. *Teaching and Teacher Education, 61*(2017), 189–198. http://doi.org/10.1016/j.tate.2016.10.015

da Ponte, J. P. (2017). Lesson studies in initial mathematics teacher education. *International Journal for Lesson and Learning Studies*, *6*(2), 169–181. https://doi.org/10.1108/IJLLS-08-2016-0021

da Ponte, J. P., Quaresma, M., & Mata-Pereira, J. (2022). Teachers' learning in lesson study: Insights provided by a modified version of the interconnected model of teacher professional growth. *ZDM Mathematics Education*, *54*(2022), 373–386. https://doi.org/10.1007/s11858-022-01367-1

Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective teacher professional development*. Learning Policy Institute. https://doi.org/10.54300/122.311

Darling-Hammond, L., Wei, R. C., Andree, A., Richardson, N., & Orphanos, S. (2009). *Professional learning in the learning profession: A status report on teacher development in the United States and abroad*. National Staff Development Council and The School Redesign Network at Stanford University. https://static1.squarespace.com/static/56b90cb101dbae64ff707585/t/583c7fe720099e25d0b1bd24/1480359912004/nsdcstudy2009.pdf

Davidoff, F., Dixon-Woods, M., Leviton, L., & Michie, S. (2015). Demystifying theory and its use in improvement. *BMJ Quality & Safety*, *24*(3), 228–238. https://doi.org/10.1136/bmjqs-2014-003627

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, *38*(3), 181–199. https://doi.org/10.3102/0013189X08331140

Dick, L. K., Appelgate, M. H., Gupta, D., & Soto, M. M. (2022). Continuous improvement lesson study: A model of MTE professional development. *Mathematics Teacher Educator*, *10*(2), 111–128 https://doi.org/10.5951/MTE.2020.0077

Doig, B., & Groves, S. (2011). Japanese lesson study: Teacher professional development through communities of inquiry. *Mathematics Teacher Education and Development*, *13*(1), 77–93.

Dolowitz, D. P. (2009). Learning by observing: Surveying the international arena. *Policy & Politics*, *37*(3), 317–334. https://doi.org/10.1332/030557309X445636

Driskell, J. E., Salas, E., & Driskell, T. (2018). Foundations of teamwork and collaboration. *American Psychologist*, *73*(4), 334–348. https://doi.org/10.1037/amp0000241

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*, *34*(2013), 107–121. https://doi.org/10.1016/j.tate.2013.04.006

Dudley, P. (2014). *Lesson Study: A Handbook*. www.lessonstudy.co.uk

Dudley, P., Xu, H., Vermunt, J. D., & Lang, J. (2019). Empirical evidence of the impact of lesson study on students' achievement, teachers' professional learning and on institutional and system evolution. *European Journal of Education*, *54*(2), 202–217. https://doi.org/10.1111/ejed.12337

Dudley, P., Pratt, M., Gilbert, C., Abbey, J., Lang, J., & Bruckdorfer, H. (2020). Cross-school 'close-to-practice' action research, system leadership and local civic partnership re-engineering an inner-city learning community. *London Review of Education*, *18*(3), 390–407. https://doi.org/10.14324/LRE.18.3.05

Elliott, J. (2019). What Is Lesson Study? *European Journal of Education*, *54*(2), 175–188. https://doi.org/10.1111/ejed.12339

Farhoush, M., Majedi, P., & Behrangi, M. (2017). Application of education management and lesson study in teaching mathematics to students of second grade of public school in district 3 of Tehran. *International Education Studies*, *10*(2), 104-113. https://doi.org/10.5539/ies.v10n2p104

Fernandez, C. (2002). Learning from Japanese approaches to professional development: The case of lesson study. *Journal of Teacher Education*, *53*(5), 393–405. https://doi.org/10.1177/002248702237394

Fernandez, C., Cannon, J., & Chokshi, S. (2003). A US–Japan lesson study collaboration reveals critical lenses for examining practice. *Teaching and Teacher Education*, *19*(2), 171–185. https://doi.org/10.1016/S0742-051X(02)00102-6

Fernandez, C., & Yoshida, M. (2004). *Lesson study: A Japanese approach to improving mathematics teaching and learning.* Lawrence Erlbaum Associates.

Festinger, L. (1962). Cognitive dissonance. *Scientific American*, *207*(4), 93–106.

Fluminhan, C., Marques, A. P., & Schünzen Junior, K. (2022). The potential of lesson study and self-efficacy to enhance teacher professional development: A systematic literature review. *TICs & EaD Em Foco*, *8*(2), 66–84. https://doi.org/10.18817/ticseademfoco.v8i2.628

Fujii, T. (2014). Implementing Japanese lesson study in foreign countries: Misconceptions revealed. *Mathematics Teacher Education and Development*, *16*(1), 65-83. https://doi.org/10.3316/aeipt.205654

Gervé, F. (2007). Lesson-Study als Modell für die schulpraktischen Studien [Lesson-Study as a model for practical school training]. In D. Flagmeyer, & M. Rotermund (Eds.). *Mehr Praxis in der Lehrerbildung – aber wie?: Möglichkeiten zur Verbesserung und Evaluation der Lehrerausbildung* [More practice in initial teacher education – but how?: Opportunities for improvement and evalution in initial teacher education] (pp. 115–130). Leipziger Universitätsverlag.

Godfrey, D., Seleznyov, S., Anders, J., Wollaston, N., & Barrera-Pedemonte, F. (2019). A developmental evaluation approach to lesson study: Exploring the impact of lesson study in London schools. *Professional Development in Education*, *45*(2), 325–340. https://doi.org/10.1080/19415257.2018.1474488

Good, T. L., Clark, S. N., & Clark, D. C. (1997). Reform efforts in American schools: Will faddism continue to impede meaningful change? In B. Biddle, T. Good, & I. Goodson (Eds.), *International handbook of teachers and teaching* (pp. 1387–1427). Springer.

Grimsæth, G., & Hallås, B. O. (2015). Lesson study model: The challenge of transforming a global idea into local practice. *Policy Futures in Education*, *14*(1), 109–122. https://doi.org/10.1177/1478210315612649

Groves, S., Doig, B., Vale, C., & Widjaja, W. (2016). Critical factors in the adaptation and implementation of Japanese Lesson Study in the Australian context. *ZDM Mathematics Education*, *48*(4), 501–512. https://doi.org/10.1007/s11858-016-0786-8

Gülhan, F. (2021). How is the Lesson Study Model Implemented in Turkey, What Does it Mean?: A Systematic Review. *Karaelmas Journal of Educational Sciences*, *9*(2021), 230–242. https://dergipark.org.tr/tr/download/article-file/2031701

Guskey, T. R. (2000). *Evaluating professional development*. Corwin Press. http://www.intuition-tlc.co.uk/wp-content/uploads/2017/08/Eval-PD-Article-Guskey.pdf

Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching*, *8*(3), 381–391. https://doi.org/10.1080/135406002100000512

Guskey, T. R. (2009). Closing the knowledge gap on effective professional development. *Educational Horizons*, *87*(4), 224–233. https://www.jstor.org/stable/42923773

Guskey, T. R. (2021). Professional learning with staying power. *Educational Leadership*, *78*(5), 54–59. https://uknowledge.uky.edu/edp_facpub/52

Guskey, T. R., & Yoon, K. S. (2009). What works in professional development? *Phi Delta Kappan*, *90*(7), 495–500. https://doi.org/10.1177%2F003172170909000709

Hadfield, M., & Jopling, M. (2016). Problematizing lesson study and its impacts: Studying a highly contextualised approach to professional learning. *Teaching and Teacher Education*, *60*(2016), 203–214. https://doi.org/10.1016/j.tate.2016.08.001

Hallitzky, M., Herfter, C., Kinoshita, E., Leicht, J., Mbaye, M., & Spenderin, K. (2021). Lesson Study in German-speaking countries: Between classroom research and teacher education. In J. Kim, N. Yoshida, S. Iwata, & H. Kawaguchi (Eds.), *Lesson Study-Based Teacher Education: The Potential of the Japanese Approach in Global Settings* (pp. 155–170). Routledge.

Hanfstingl, B., Rauch, F., & Zehetmeier, S. (2019). Lesson study, learning study and action research: Are there more differences than a discussion about terms and schools? *Educational Action Research*, *27*(4), 455–459. https://doi.org/10.1080/09650792.2019.1652450

Harland, J., & Kinder, K. (2014). Teachers' continuing professional development: Framing a model of outcomes. *Professional Development in Education*, *40*(4), 669–682. https://doi.org/10.1080/19415257.2014.952094

Harmon-Jones, E., & Mills, J. (2019). An introduction to cognitive dissonance theory and an overview of current perspectives on the theory. In E. Harmon-Jones (Eds.), *Cognitive dissonance: Reexamining a pivotal theory in psychology (2nd ed.*, pp. 3–24). American Psychological Association. https://doi.org/10.1037/0000135-001

Helmke, A., & Helmke, T. (2015). Unterrichtsdiagnostik als Ausgangspunkt für Unterrichtsentwicklung. [Teaching diagnostics as a starting point for teaching development]. In H.-G. Rolff (Ed.), *Handbuch Unterrichtsentwicklung* [*Handbook for the Development of Instructions*](pp. 242–257). Beltz.

Hervas, G. (2021a). Lesson study as a faculty development initiative in higher education: A systematic review. *AERA Open*, *7*(1), 1–19. https://doi.org/10.1177/2332858420982564

Hervas, G. (2021b). The international popularisation of lesson study: Early studies and their relevance in later literature. *Revista Panamericana de Pedagogía*, *33*(2022), 38–45. https://doi.org/10.21555/rpp.v33i33.2397

Hiebert, J., & Stigler, J. W. (2000). A Proposal for Improving Classroom Teaching: Lessons from the TIMSS Video Study. *Elementary School Journal*, *101*(1), 3–20. https://doi.org/10.1086/499656

Hiebert, J., Gallimore, R., & Stigler, J. W. (2002). A knowledge base for the teaching profession: What would it look like and how can we get one? *Educational Researcher*, *31*(5), 3–15. https://doi.org/10.3102/0013189X031005003

Hinojosa, A. S., Gardner, W. L., Walker, H. J., Cogliser, C., & Gullifor, D. (2017). A review of cognitive dissonance theory in management research: Opportunities for further development. *Journal of Management*, *43*(1), 170–199. https://doi.org/10.1177/0149206316668236

Hrastinski, S. (2021). Digital tools to support teacher professional development in lesson studies: A systematic literature review. *International Journal for Lesson & Learning Studies*, *10*(2), 138–149. https://doi.org/10.1108/IJLLS-09-2020-0062

Huang, R., & 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*, *48*(4), 393–409. https://doi.org/10.1007/s11858-016-0795-7

Hunzicker, J. (2011). Effective professional development for teachers: A checklist. *Professional Development in Education*, *37*(2), 177–179. https://doi.org/10.1080/19415257.2010.523955

Hwang, Y. R., & Fwu, B.-J. (2011). The most significant factor underlying Japanese students' academic success: Lesson study. *Journal of Research in Education Sciences*, *56*(4), 69–97. https://doi.org/10.3966/2073753X2011125604003

Igira, F. T., & Gregory, J. (2009). Cultural historical activity theory. In Y. K. Dwievedi, B. Lai, M. D. Williams, S. L. Schneberger, & M. Wade (Eds.), *Handbook of research on contemporary theoretical models in information systems* (pp. 434–454). Igi Global.

Ilgen, D. R., Hollenbeck, J. R., Johnson, M., & Jundt, D. (2005). Teams in organizations: From input-process-output models to IMOI models. *Annual Review of Psychology*, *56*(1), 517–543. https://doi.org/10.1146/annurev.psych.56.091103.070250

Jakobeit, M., Westphal, A., Jurczok, A., Kager, K., & Vock, M. (2021). Welche schulischen Rahmenbedingungen sind für die Kooperation von Lehrkräften in Lesson-Study-Prozessen förderlich? Eine Clusteranalyse von Grundschulen im Rahmen des Forschungs- und Schulentwicklungsprojekts Leistung macht Schule (LemaS) [What educational conditions are beneficial for teachers' collaboration in lesson-study projects? In the context of the research and school development project Leistung macht Schule, a cluster analysis of primary schools was conducted]. *Journal Für Schulentwicklung*, *2*(21), 17–26. https://www.mathtransformations.com/_files/ugd/e8bafa_519d47a1ee3544a3b6627006488fb271.pdf#page=18

Johnson, D. W., & Johnson, R. T. (2009). Energizing learning: The instructional power of conflict. *Educational Researcher*, *38*(1), 37–51. https://doi.org/10.3102/0013189X08330540

Joubert, J., Callaghan, R., & Engelbrecht, J. (2020). Lesson study in a blended approach to support isolated teachers in teaching with technology. *ZDM: The International*

Journal on Mathematics Education, *52*(5), 907–925. https://doi.org/10.1007/s11858-020-01161-x

Jurczok, A., Kalinowski, E., Zaruba, N., & Vock, M. (2020). Teilprojekt 22. Kooperative Unterrichtsentwicklung durch die Lesson Study-Methode. [Cooperative lesson development through the lesson study method]. In G. Weigand, M. Vock, F. Preckel, C. Fischer, C. Perleth, F. Käpnick, & H.-W. Wollersheim (Eds.), *Leistung macht Schule–Konzepte für die Förderung begabter und besonders leistungsfähiger Schülerinnen und Schüler* [Excellence in School Education–*Strategies to faciltiate high-achiving and potentially high-achieving students*](pp. 224–234). Beltz.

Kager, K., Kalinowski, E., Jurczok, A., & Vock, M. (2021). *A systematic review of methodological transparency in lesson study: How do we report on the observation and reflection stages?* OSF. https://doi.org/10.17605/OSF.IO/5NXGY

Kager, K., Bolli, S., Jurczok, A., & Vock, M. (2022). Lernen durch Konflikt? Ein praktischer Erfahrungsbericht über kooperative Unterrichtsentwicklung durch Lesson Study [Learning through conflict? An experience report on the collaborative development of instructions through lesson study] In G. Weigand, C. Fischer, F. Käpnick, C. Perleth, F. Preckel, M. Vock, & H.-W. Wollersheim, *Dimensionen der Begabungs- und Begabtenförderung in der Schule* [Dimensions of promoting gifted education in schools](Vol. 2, pp. 115–128). wbv. https://doi.org/10.3278/9783763967858

Kalinowski, E., Jurczok, A., & Vock, M. (2021). Mit der Lesson Study-Methode zu einem leistungsfördernden Unterricht in der Grundschule? [With the Lesson Study-Method to a performance-promoting lesson in elementary school?] In B. Klopsch & A. Sliwka (Eds.), *Kooperative Professionalität: Internationale Ansätze der ko-konstruktiven Unterrichtsentwicklung* [Cooperative Professionality: Global Approaches to Co-Constructive Education Development](pp. 138–153). Beltz Juventa.

Kanellopoulou, E.-M. D., & Darra, M. (2019). Benefits, difficulties and conditions of lesson study implementation in basic teacher education: A review. *International Journal of Higher Education*, *8*(4), 18–35. https://doi.orf/10.5430/ijhe.v8n4p18

Kang, S., Scharmann, L. C., & Noh, T. (2004). Reexamining the role of cognitive conflict in science concept learning. *Research in Science Education*, *34*(1), 71–96. https://doi.org/10.1023/B:RISE.0000021001.77568.b3

Karlsen, A. M. F., & Helgevold, N. (2019). Lesson study: Analytic stance and depth of noticing in post-lesson discussions. *International Journal for Lesson and Learning Studies*, *8*(4), 290–304. https://doi.org/10.1108/IJLLS-04-2019-0034

Kawaguchi, H., & Iwata, S. (2021). The potential of lesson study-based teacher education. In J. Kim, N. Yoshida, S. Iwata, & H. Kawaguchi (Eds.), *Lesson Study-based teacher education: The potential of the Japanese approach in global settings* (pp. 53–68). Routledge.

Kennedy, M. M. (2016). How does professional development improve teaching? *Review of Educational Research*, *86*(4), 945–980. https://doi.org/10.3102/0034654315626800

Kim, J. (2021). Through foreign eyes: A critical understanding of lesson study-based teacher education in Japan. In J. Kim, N. Yoshida, S. Iwata, & H. Kawaguchi (Eds.), *Lesson Study-based Teacher Education* (pp. 9–28). Routledge.

Kincal, R. Y., Ozan, C., & Ileritürk, D. (2019). Increasing students' English language learning levels via lesson study. *English Language Teaching*, *12*(9), 88–95. https://doi.org/10.5539/elt.v12n9p88

Klopsch, B., & Sliwka, A. (2021). Kooperative Professionalität als internationaler Ansatz der Unterrichtsentwicklung [Cooperative professionality as an international approach to

the development of instructions]. In B. Klopsch & A. Sliwka (Eds.), *Kooperative Professionalität: Internationale Ansätze der ko-konstruktiven Unterrichtsentwicklung* [Cooperative professionality: International approaches to the co-constructive developement of instructions] (pp. 7–19). Beltz Juventa.

Knapp, A., Bomer, M., & Moore, C. (2011). Lesson study as a learning environment for coaches of mathematics teachers. In L. C. Hart, A. S. Alston, & A. Murata (Eds.), *Lesson study research and practice in mathematics education* (pp. 153–164). Springer. https://doi.org/10.1007/978-90-481-9941-9_12

Knoblauch, R. (2017). Lesson Study: Kooperative Weiterentwicklung des Lehrens und Lernens [Lesson study: Cooperative professionalization of teaching and learning], *Pädagogik*, 3, 34–39.

Korthagen, F. (2016). Inconvenient truths about teacher learning: Towards professional development 3.0. *Teachers and Teaching, 23*(4) 387–405. https://doi.org/10.1080/13540602.2016.1211523

Korthagen, F. A. J. (1985). Reflective teaching and preservice teacher education in the Netherlands. *Journal of Teacher Education*, *36*(5), 11–15. https://doi.org/10.1177/002248718503600502

Korthagen, F., & Vasalos, A. (2005). Levels in reflection: Core reflection as a means to enhance professional growth. *Teachers and Teaching*, *11*(1), 47–71. https://doi.org/10.1080/1354060042000337093

Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013). Single-case intervention research design standards. *Remedial and Special Education*, *34*(1), 26–38. https://doi.org/10.1177/0741932512452794

Kullmann, H., & Friedli, R. (2012). Lesson Study: Eine Methode zur symbiotischen Entwicklung von Unterricht und Lehrerprofessionalität durch Praxisforschung in der dritten Phase der Lehrerbildung [Lesson study: A method for the symbiotic development of instructions and teachers' professionality through practical research in the third phase of initial teacher education]. In C. Freitag & I. Bargen (Eds.). *Praxisforschung in der Lehrerbildung: 16. Jahrestagung des Nordverbundes Schulbegleitforschung am 15./16. September 2011 an der Universität Paderborn* [Practical research in initial teacher education: 16th Annual conference of the Nordverbund Schulbegeleitungsforschung on the 15th/ 16th September 2011 at the University of Paderborn](pp. 133–144). LIT Verlag.

Larssen, D. L. S., Cajkler, W., Mosvold, R., Bjuland, R., Helgevold, N., Fauskanger, J., Wood, P., Baldry, F., Jakobsen, A., Bugge, H. E., Næsheim-Bjørkvik, G., & Norton, J. (2018). A literature review of lesson study in initial teacher education: Perspectives about learning and observation. *International Journal for Lesson and Learning Studies*, *7*(1), 8–22. https://doi.org/10.1108/IJLLS-06-2017-0030

Lee, L. H. J., & Tan, S. C. (2020). Teacher learning in lesson study: Affordances, disturbances, contradictions, and implications. *Teaching and Teacher Education*, *89(2020)*, 1–15. https://doi.org/10.1016/j.tate.2019.102986

Lemire, S., Christie, C. A., & Inkelas, M. (2017). The methods and tools of improvement science: The methods and tools of improvement science. *New Directions for Evaluation*, *153*(2017), 23–33. https://doi.org/10.1002/ev.20235

Lewis, C. (2000). *Lesson Study: The core of Japanese professional development* [Conference Paper]. Annual Meeting of the American Educational Research Association, New Orleans, LA.

Lewis, C. (2002). *Lesson study: A handbook of teacher-led instructional change.* Research for Better Schools.

Lewis, C. (2009). What is the nature of knowledge development in lesson study? *Educational Action Research*, *17*(1), 95–110. https://doi.org/10.1080/09650790802667477

Lewis, C. (2015). What is improvement science? Do we need it in education? *Educational Researcher*, *44*(1), 54–61. https://doi.org/10.3102/0013189X15570388

Lewis, C. (2016). How does lesson study improve mathematics instruction? *ZDM Mathematics Education*, *48*(4), 571–580. https://doi.org/10.1007/s11858-016-0792-x

Lewis, C., Friedkin, S., Emerson, K., Henn, L., & Goldsmith, L. (2019). How does lesson study work? Toward a theory of lesson study process and impact. In R. Huang, A. Takahashi, & J. P. da Ponte (Eds.), *Theory and Practice of Lesson Study in Mathematics* (pp. 13–37). Springer International Publishing. https://doi.org/10.1007/978-3-030-04031-4_2

Lewis, C., & Perry, R. (2014). Lesson Study with Mathematical Resources: A Sustainable Model for Locally-Led Teacher Professional Learning. *Mathematics Teacher Education and Development*, *16*(1), 22–42. https://doi.org/10.3316/aeipt.205652

Lewis, C., & Perry, R. (2017). Lesson study to scale up research-based knowledge: A randomized, controlled trial of fractions learning. *Journal for Research in Mathematics Education*, *48*(3), 261–299. https://doi.org/10.5951/jresematheduc.48.3.0261

Lewis, C., Perry, R., & Friedkin, S. (2009). Lesson study as action research. In B. Somekh & S. Noffke (Eds.), *The SAGE handbook of educational action research* (pp. 142–154). Sage.

Lewis, C., Perry, R., & Murata, A. (2006). How should research contribute to instructional improvement? The case of lesson study. *Educational Researcher*, *35*(3), 3–14. https://doi.org/10.3102/0013189X035003003

Lewis, C., Perry, R., & Hurd, J. (2005). *Lesson study: A theoretical model and a North American case.* [Manuscript submitted for publication].

Limon, M. (2001). On the cognitive conflict as an instructional strategy for conceptual change: A critical appraisal. *Learning and Instruction*, *11*(4–5), 357–380. https://doi.org/10.1016/S0959-4752(00)00037-2

Lipowsky, F., & Rzejak, D. (2015). Key features of effective professional development programmes for teachers. *Ricercazione*, *7*(2), 27–51.

Maddux, C. (2003). Fads, distance education, and the importance of theory. *Computers in the Schools*, *20*(3), 121–127. https://doi.org/10.1300/J025v20n03_11

Maddux, C., & Cummings, R. (2004). Fad, Fashion, and the Weak Role of Theory and Research in Information Technology in Education. *Journal of Technology and Teacher Education*, *12*(4), 511–533. https://link.gale.com/apps/doc/A126331245/AONE?u=anon~c51e6067&sid=googleScholar&xid=0dd8bddf

Marks, R., Bertram, S., & Eilks, I. (2008). Learning chemistry and beyond with a lesson plan on potato crisps, which follows a socio-critical and problem-oriented approach to chemistry lessons—A case study. *Chemistry Education Research and Practice*, *9*(3), 267–276. https://doi.org/10.1039/B812416G

Massenkeil, J., & Rothland, M. (2016). Kollegiale Kooperation im Lehrerberuf. Überblick und Systematisierung aktueller Forschung. [Collegial cooperation in the teaching profession. Overview and systematization of current research]. *Schulpädagogik heute*, *7*(13), 1-28.

Mathieu, J. E., Gallagher, P. T., Domingo, M. A., & Klock, E. A. (2019). Embracing complexity: Reviewing the past decade of team effectiveness research. *Annual Review of Organizational Psychology and Organizational Behavior*, *6*(1), 17–46. https://doi.org/10.1146/annurev-orgpsych-012218-015106

Mayring, P. (2010). Qualitative Inhaltsanalyse [Qualitative content analysis]. In G. Mey & K. Mruck (Eds.), *Handbuch qualitative Forschung in der Psychologie* [*Handbook of qualitative research in psychology*](pp. 601-613). Wiesbaden: VS Verlag für Sozialwissenschaften.

Mercer, N. (2004). Sociocultural discourse analysis. *Journal of Applied Linguistics*, *1*(2), 137–168. https://doi.org/doi:10.1558/ japl.2004.1.2.137

Mezirow, J. (1990). *How critical reflection triggers transformative learning*. Jossey-Bass Publishers.

Mezirow, J. (2000). *Learning as transformation: Critical perspectives on a theory in progress*. Jossey-Bass Publishers.

Moravcsik, A. (2020). *Transparency in qualitative research*. SAGE Publications Ltd. https://doi.org/10.4135/9781526421036

Moss, J., Hawes, Z., Naqvi, S., & Caswell, B. (2015). Adapting Japanese lesson study to enhance the teaching and learning of geometry and spatial reasoning in early years classrooms: A case study. *ZDM: The International Journal on Mathematics Education*, *47*(3), 377–390. https://doi.org/10.1007/s11858-015-0679-2

Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology*, *18*(143), 1–7. https://doi.org/10.1186/s12874-018-0611-x

Murase, M. (2007). Jugyo kenkyuu no genzai [State-of-the-art of lesson study research]. *Kyouikugaku Kenkyu* [The Japanese Journal of Educational Research], *74*(1), 41–48.

Murata, A. (2004). Paths to Learning Ten-Structured Understandings of Teen Sums: Addition Solution Methods of Japanese Grade 1 Students. *Cognition and Instruction*, *22*(2), 185–218. https://doi.org/10.1207/s1532690xci2202_2

Murata, A. (2011). Introduction: Conceptual overview of lesson study. In L. C. Hart, A. S. Alston, & A. Murata (Eds.), *Lesson Study Research and Practice in Mathematics Education* (pp. 1–12). Springer Netherlands. https://doi.org/10.1007/978-90-481-9941-9_1

Murphy, R., Weinhardt, F., Wyness, G., & Rolfe, H. (2017). Lesson study: Evaluation report and executive summary. *Education Endowment Foundation*. https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/Lesson_Study.pdf

Murray, F. B. (1983). Equilibration as cognitive conflict. *Developmental Review*, *3*(1), 54–61. https://doi.org/10.1016/0273-2297(83)90007-2

Mynott, J. (2017). A primary head teacher's exploration of lesson study. [Doctoral dissertation, University of Hertfordshire]. University of Hertfordshire Research Archive. https://uhra.herts.ac.uk/handle/2299/18330

Mynott, J. (2019). Lesson study outcomes: A theoretical model. *International Journal for Lesson and Learning Studies*, *8*(2), 117–134. https://doi.org/10.1108/IJLLS-08-2018-0057

Nijhawan, S. (2017). Bridging the gap between theory and practice with design-based action research. *Studia Paedagogica*, *22*(4), 9–29. https://doi.org/10.5817/SP2017-4-2

Norwich, B. (2015). Educational psychology, neuroscience and lesson study: translating research knowledge into practice requires teacher research. *Knowledge Cultures, (3)*2, 7-25. https://www.ceeol.com/search/article-detail?id=413176

Norwich, B. (2018). Making sense of international variations in lesson study and lesson study-like practices: An exploratory and conceptual perspective. *International Journal for Lesson and Learning Studies*, *7*(3), 201–216. https://doi.org/10.1108/IJLLS-02-2018-0007

Norwich, B., Benham-Clarke, S., & Goei, S. L. (2021). Review of research literature about the use of lesson study and lesson study-related practices relevant to the field of special needs and inclusive education. *European Journal of Special Needs Education*, *36*(3), 309–328. https://doi.org/10.1080/08856257.2020.1755929

Norwich, B., Koutsouris, G., Fujita, T., Ralph, T., Adlam, A., & Milton, F. (2016). Exploring knowledge bridging and translation in lesson study using an inter-professional team. *International Journal for Lesson and Learning Studies*, *5*(3), 180–195. https://doi.org/10.1108/IJLLS-02-2016-0006

Norwich, B., & 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*, *34*(2013), 162–173. https://doi.org/10.1016/j.tate.2013.04.012

Norwich, B., & Ylonen, A. (2015). A design-based trial of lesson study for assessment purposes: Evaluating a new classroom based dynamic assessment approach. *European Journal of Special Needs Education*, *30*(2), 253–273. https://doi.org/10.1080/08856257.2015.1009702

Ono, Y. (2009). Kyoiku program no borrowing and lending [Borrowing and lending of educational programs]. *Kokusai Kyoiku Kyoryoku Ronsyu [Journal of International Cooperation in Education],12*(2), 69–80.

Opfer, V. D., & Pedder, D. (2011). Conceptualizing teacher professional learning. *Review of Educational Research*, *81*(3), 376–407. https://doi.org/10.3102/0034654311413609

Park, S., Hironaka, S., Carver, P., & Nordstrum, L. (2013). Continuous improvement in education. Advancing teaching-improving learning. *Carnegie Foundation for the Advancement of Teaching*. https://files.eric.ed.gov/fulltext/ED560145.pdf

Parks, A. N. (2008). Messy learning: Preservice teachers' lesson-study conversations about mathematics and students. *Teaching and Teacher Education*, *24*(5), 1200–1216. https://doi.org/10.1016/j.tate.2007.04.003

Pérez Granados, L., Alcaraz-Salarirche, N., Fernández Navas, M., & Postigo Fuentes, A. Y. (2022). Action research through lesson study: A space for learning in initial teacher training. *Educational Action Research*, 1–18. https://doi.org/10.1080/09650792.2022.2108473

Pogrow, S. (2017). The failure of the U.S. education research establishment to identify effective practices: Beware effective practices policies. *Education Policy Analysis Archives*, *25*(5), 1–22. https://doi.org/10.14507/epaa.25.2517

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, *66*(2), 211–227. https://doi.org/10.1002/sce.3730660207.

Quaresma, M., & da Ponte, J. P. (2019). Primary teachers' reflection and collaboration dynamics in a mathematics lesson study. *Bolema - Mathematics Education Bulletin*, *33*(63), 368–388. https://doi.org/10.1590/1980-4415v33n63a18

6. References

Richit, A., & da Ponte, J. P. (2019). Professional collaboration in lesson studies in the perspective of participants teachers. *Bolema - Mathematics Education Bulletin*, *33*(64), 937–962. https://doi.org/10.1590/1980-4415v33n64a24

Richter, D., & Pant, H. A. (2016). *Lehrerkooperation in Deutschland*. *Eine Studie zu kooperativen Arbeitsbeziehungen bei Lehrkräften der Sekundarstufe [Collaboration of teachers in Germany. A study on the cooperative work relationships of teachers in highschool]*. Bertelsmann Stiftung. https://www.telekom-stiftung.de/sites/default/files/files/media/publications/BST%20Lehrerkooperation%20in%20Deutschland.pdf

Richter, E., & Richter, D. (2020). Fort- und Weiterbildung von Lehrpersonen [Professional development for teachers]. In C. Cramer, J. König, M. Rothland, & S. Blömeke, *Handbuch Lehrerinnen- und Lehrerbildung [Handbook for teacher training]*. Verlag Julius Klinkhardt. https://doi.org/10.35468/hblb2020-040

Rodgers, C. (2002). Defining reflection: Another look at John Dewey and reflective thinking. *Teachers College Record*, *104*(4), 842–866. https://doi.org/10.1111/1467-9620.00181

Rolff, H.-G. (2019). *Schulentwicklung auf den Punkt gebracht* [School development put in a nutshell]. Debus Pädagogik.

Rzejak, D. (2019). Zur Wirksamkeit von Lesson Study. [On the effectiveness of lesson study]. In C. Mewald & E. Rauscher (Eds.), *Lesson Study. Das Handbuch für kollaborative Unterrichtsentwicklung und Lernforschung [Handbook on the collaborative development and exploration of instructions]* (pp. 97–111). Studien Verlag.

Saito, E. (2012). Key issues of lesson study in Japan and the United States: A literature review. *Professional Development in Education*, *38*(5), 777–789. https://doi.org/10.1080/19415257.2012.668857

Saito, E., Khong, T. D. H., Hidayat, A., Hendayana, S., & Imansyah, H. (2020). Typologies of lesson study coordination: A comparative institutional analysis. *Professional Development in Education*, *46*(1), 65–81. https://doi.org/10.1080/19415257.2018.1561495

Sannino, A., & Engeström, Y. (2018). Cultural-historical activity theory: Founding insights and new challenges. *Cultural-Historical Psychology*, *14*(3), 43–56. https://doi.org/10.17759/chp.2018140305

Schipper, T., Goei, S. L., de Vries, S., & van Veen, K. (2017). Professional growth in adaptive teaching competence as a result of lesson study. *Teaching and Teacher Education*, *68*(2017), 289–303. https://doi.org/10.1016/j.tate.2017.09.015

Schipper, T., Goei, S. L., de Vries, S., & van Veen, K. (2018). Developing teachers' self-efficacy and adaptive teaching behaviour through lesson study. *International Journal of Educational Research*, *88*(2018), 109–120. https://doi.org/10.1016/j.ijer.2018.01.011

Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. Basic Books.

Schön, D. A. (1995). *The reflective practitioner: How professionals think in action*. Ashgate.

Schreier, M. (2012). *Qualitative content analysis in practice.* New York: Sage publications.

Seleznyov, S. (2018). Lesson study: An exploration of its translation beyond Japan. *International Journal for Lesson and Learning Studies*, *7*(3), 217–229. https://doi.org/10.1108/IJLLS-04-2018-0020

Seleznyov, S. (2019). Lesson study beyond Japan: Evaluating impact. *International Journal for Lesson and Learning Studies*, *8*(1), 2–18. https://doi.org/10.1108/IJLLS-09-2018-0061

Seleznyov, S., Goei, S. L., & Ehren, M. (2021). International policy borrowing and the case of Japanese lesson study: Culture and its impact on implementation and adaptation.

Professional Development in Education, 1–15. https://doi.org/10.1080/19415257.2021.1973069

Shanmugam, S. K. S., Chinnappan, M., & Leong, K. E. (2020). Action research in examining the enquiry approach of lesson study in mathematics. *Pertanika J. Soc. Sci. & Hum.*, *28*(3), 1675–1693. http://www.pertanika.upm.edu.my/resources/files/Pertanika%20PAPERS/JSSH%20Vol.%2028%20(3)%20Sep.%202020/05%20JSSH-4813-2019.pdf

Shu, L. L., Gino, F., & Bazerman, M. H. (2011). Dishonest deed, clear conscience: When cheating leads to moral disengagement and motivated forgetting. *Personality and Social Psychology Bulletin*, *37*(3), 330–349. https://doi.org/10.1177/0146167211398138

Slavin, R. E. (1999). Chapter 15—The pendulum revisited: Faddism in education and its alternatives. In G. J. Cizek (Eds.), *Handbook of Educational Policy* (pp. 373–386). Academic Press. https://doi.org/10.1016/B978-012174698-8/50042-4

Slavit, D., & Nelson, T. H. (2010). Collaborative teacher inquiry as a tool for building theory on the development and use of rich mathematical tasks. *Journal of Mathematics Teacher Education*, *13*(3), 201–221. https://doi.org/10.1007/s10857-009-9136-x

Somekh, B., & Zeichner, K. (2009). Action research for educational reform: Remodelling action research theories and practices in local contexts. *Educational Action Research*, *17*(1), 5–21. https://doi.org/10.1080/09650790802667402

Soto Gómez, E., Serván Núñez, M. J., Trapero, N. P., & Pérez Gómez, Á. I. (2019). Action research through lesson study for the reconstruction of teachers' practical knowledge. A review of research at Málaga University (Spain). *Educational Action Research*, *27*(4), 527–542. https://doi.org/10.1080/09650792.2019.1610020

Stigler, J. W., & Hiebert, J. (1999). *The teaching gap: Best ideas from the world's teachers for improving education in the classroom*. Summit Books.

Stigler, J. W., & Hiebert, J. (2016). Lesson study, improvement, and the importing of cultural routines. *ZDM: The International Journal on Mathematics Education*, *48*(4), 581–587. https://doi.org/10.1007/s11858-016-0787-7

Stufflebeam, D. L. (1983). The CIPP model for program evaluation. In D. L. Stufflebeam, G. Madaus, & T. Kellaghan (Eds.), *Evaluation models: Viewpoints on Educational and Human Services Evaluation* (pp. 117–141). Springer. https://doi.org/10.1007/978-94-009-6669-7_7

Stylianou, P., & Zembylas, M. (2019). Ethical and political dimensions of action research and lesson study: Reflections from a research project on a controversial issue in Cyprus. *Educational Action Research*, *27*(4), 581–594. https://doi.org/10.1080/09650792.2018.1483835

Takahashi, A. (2014). The Role of the Knowledgeable Other in Lesson Study: Examining the Final Comments of Experienced Lesson Study Practitioners. *Mathematics Teacher Education and Development*, *16*(1), 4–21. https://doi.org/10.3316/aeipt.205651

Takahashi, A., & McDougal, T. (2016). Collaborative lesson research: Maximizing the impact of lesson study. *ZDM Mathematics Education*, *48*(4), 513–526. https://doi.org/10.1007/s11858-015-0752-x

Tidd, J., & Bessant, J. (2018). Innovation management challenges: From fads to fundamentals. *International Journal of Innovation Management*, *22*(05), 1–14. https://doi.org/10.1142/S1363919618400078

6. References

Tinoca, L., Piedade, J., Santos, S., Pedro, A., & Gomes, S. (2022). Design-based research in the educational field: A systematic literature review. *Education Sciences*, *12*(6), 1–19 https://doi.org/10.3390/educsci12060410

Uştuk, Ö., & Çomoğlu, İ. (2019). Lesson study for professional development of English language teachers: Key takeaways from international practices. *Journal on Efficiency and Responsibility in Education and Science*, *12*(2), 41–50. https://doi.org/10.7160/eriesj.2019.120202

van Es, E., & Sherin, M. (2002). Learning to notice: Scaffolding new teachers' interpretations of classroom interactions. *Journal of Information Technology for Teacher Education*, *10*(4), 571–596. https://tinyurl.com/yc82twtp

VERBI Software. (2019). *MaxQda 2020*. VERBI Software. https://www.maxqda.com

Vosniadou, S., Lawson, M. J., Wyra, M., Van Deur, P., Jeffries, D., & I Gusti Ngurah, D. (2020). Pre-service teachers' beliefs about learning and teaching and about the self-regulation of learning: A conceptual change perspective. *International Journal of Educational Research*, *99(2020)*, 1–19. https://doi.org/10.1016/j.ijer.2019.101495

Vygotsky, L. S. (1986). *Thought and language.* Massachusetts Institute of Technology.

Waights, S. (2014). *Guide to scoring methods using the Maryland Scientific Methods Scale*. What Works center for local economic growth [Website]. https://whatworksgrowth.org/resource-library/the-maryland-scientific-methods-scale-sms/

Wang, H., Gibson, C., & Zander, U. (2020). Editors' comments: Is research on corporate social responsibility undertheorized? *Academy of Management Review*, *45*(1), 1–6. https://doi.org/10.5465/amr.2019.0450

Warwick, P., Vrikki, M., Vermunt, J. D., Mercer, N., & 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*, *48*(4), 555–569. https://doi.org/10.1007/s11858-015-0750-z

Watanabe, T., Takahashi, A., & Barham, A. I. (2019). Implementing school-wide collaborative lesson research in Qatar. *Journal of Institutional Research South East Asia*, *17*(2), 47–70. http://hdl.handle.net/10576/14352

Wei, G., & Huang, R. (2022). Research-practice partnerships in lesson and learning studies: A review from Asian experiences. *Asia Pacific Journal of Education*, *42*(1), 138–153. https://doi.org/10.1080/02188791.2022.2031876

Weigand, G. (2022). Einleitung [Introduction]. In G. Weigand, C. Fischer, F. Käpnick, C. Perleth, F. Preckel, M. Vock, M., & H.-W.Wollersheim (pp. 9-16). *Dimensionen der Begabungs- und Begabtenförderung in der Schule* [Dimensions of promoting gifted education in schools]. wbv.

Widjaja, W., Vale, C., Groves, S., & Doig, B. (2017). Teachers' professional growth through engagement with lesson study. *Journal of Mathematics Teacher Education*, *20*(4), 357–383. https://doi.org/10.1007/s10857-015-9341-8

Willems, I., & van den Bossche, P. (2019). Lesson study effectiveness for teachers' professional learning: A best evidence synthesis. *International Journal for Lesson and Learning Studies*, *8*(4), 257–271. https://doi.org/10.1108/IJLLS-04-2019-0031

Willis, J. W., & Edwards, C. (2014). *Action research: Models, methods, and examples*. IAP.

Xu, H., & Pedder, D. (2014). Lesson Study: An international review of the research. In P. Dudley (Ed.), *Lesson Study: Professional Learning for Our Time* (pp. 29–59). https://doi.org/10.4324/9780203795538

Ylonen, A., & Norwich, B. (2013). Professional learning of teachers through a lesson study process in England: Contexts, mechanisms and outcomes. *International Journal for Lesson and Learning Studies*, *2*(2), 137–154. https://doi.org/10.1108/20468251311323388

Yoon, K. S., Duncan, T., Wen-Yu Lee, S., Scarloss, & Shapley, K. (2007). *Reviewing the evidence on how teacher professional development affects student achievement (Issues & Answers Report, REL 2007–No. 033)* (p. 62). Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. http://ies.ed.gov/ncee/edlabs

Yoshida, M. (1999). *Lesson study: A case study of a Japanese approach to improving instruction through school-based teacher development [doctoral dissertation]*. The University of Chicago.

Yoshida, N., Matsuda, M., & Miyamoto, Y. (2021). The landscape of lesson study: A methodology for teachers' professional development and educational research. In J. Kim, N. Yoshida, S. Iwata, & H. Kawaguchi (Eds.), *Lesson Study-Based Teacher Education: The Potential of the Japanese Approach in Global Settings* (pp. 29–50). Routledge.

Yoshida, N., Matsuda, M., & Miyamoto, Y. (2021). Intercultural collaborative lesson study between Japan and Germany. *International Journal for Lesson & Learning Studies*, *10*(3), 245–259. https://doi.org/10.1108/IJLLS-07-2020-0045

Yurkofsky, M. M., Peterson, A. J., Mehta, J. D., Horwitz-Willis, R., & Frumin, K. M. (2020). Research on continuous improvement: Exploring the complexities of managing educational change. *Review of Research in Education*, *44*(1), 403–433. https://doi.org/10.3102/0091732X20907363

7. Appendix

Table A1

Overview of Literature Reviews and Systematic Reviews on Lesson Study and their Main Findings and Implications

study	year	type	topic & research aims	time	studies included	main findings and implications
Saito	2012	literature review	current state-of-the art in the research field, progress and limitations of LS research in the US and Japan	all time	27	- the observation and reflection stages are often reported as difficult to implement as not all teachers possess the necessary skills - Lesson Study for the Learning Community (LSLC) is of systemic nature and can enhance LS as implemented in schools, as school leaders are included in LS and problems on a school level are addressed - educators and researchers in Japan should engage more in the international discourse on LS - the field needs more research into the stages of observation and reflection
Cheung & Wong	2014	review and systematic review	existent evidence for the impact of LS and Learning Studies on teachers and students	2000 - 2010	20 (review) + 9 (syste-matic review)	- the majority of studies reported a positive impact of LS on students and teachers - the reviewed studies employed different outcome measures and research designs (6 case studies, 2 quasi-experimental trials, 1 controlled experimental trial) - some studies lacked information (e.g., number of teachers or students that participated in a LS cycle) - the evidence for the positive impact of LS on teachers and students was not conclusive as the field is lacking "high-quality and well-controlled studies" - in order to be able to conduct a meta-analysis, LS research needs to provide more detailed information, use similar methodologies and outcome measures
Xu & Pedder	2014	systematic review	growth and geographical spread of LS research	2002 - 2013	67	- geographical distribution: majority of articles on LS have been published in North America (34) and Asia (23) - school setting: majority of studies are carried out in primary schools, secondary schools, and initial teacher education - subject focus: majority of studies focus on mathematics - sample characteristics: studies varied in sample size, settings adaptations, and variations - research focus and findings: 73% of articles focused on benefits and constraints that influence LS, 14% focused on how LS is used by teachers to examine aspects of teaching and learning, 7% of articles examined how mechanisms of LS and how teachers learn through LS, 6% focused on contextual features and how LS can be implemented successfully - only few of the review articles provide any or sufficient theorization of teacher learning - the majority of studies provide weak explanatory power, as they are small-scale, qualitative, and exploratory in nature - the evidence collected in the majority of studies is predominantly not rigorous and trustworthy - research areas neglected in the reviewed studies: teacher talk, collaboration, and aspects such as trust, conflict, collegiality - there is a lack of studies that investigate the mechanisms of LS and the processes through which teachers learn
Huang & Shimizu	2016	literature review	synthesis of the state-of-the-art of research on LS with in-service mathematics teachers	all time	52	- the spread of LS outside of Japan and China tends to be project- or research-based, while LS in Japan and China has evolved as a fixed part of teacher education over time - LS is used for different purposes in different countries, these purposes shape the LS implementations and the power-relations within the LS team

						- the strategies of how teachers focus on students' learning and differ between countries (Japan: anticipating student learning, collecting student work; Hong-Kong and Sweden: focus on interviews and pre-post-tests, UK: focuses on case students) - the reviewed studies that clearly connected LS to theory either use cognitive theories or socio-cultural theory
da Ponte	2017	literature review	adaptations, aims, and outcomes of LS in initial teacher education (prospective mathematics teachers)	all time	16	- majority of reviewed studies was carried out in the US - the reviewed studies showed a high variability in LS adaption and design, in their theoretical frameworks, and in the manner that these frameworks were used - some of the reviewed studies did not reference any theoretical framework - the reviewed studies used different outcome measures and defined the purpose of LS and its outcomes in various ways - the field needs further research into the relationship between students and instructors in LS teams in initial teacher education, into the problem of scale, and how LS is adapted and simplified
Larsson et al.	2018	literature review	how learning and observation are discussed in studies on LS in initial teacher education	all time	24	- the majority of LS research focuses on LS with in-service teachers - the majority of reviewed studies were conducted in the US, followed by Europe - 11 studies reported on learning in a coherent manner and discussed their findings in the light of the theoretical framework, 3 studies reported on learning in a partially coherent manner, and 11 studies made little use of theoretical frameworks and did not discuss their findings in relation to a theory - majority of reviewed studies only described in passing how teachers observed, some papers did not provide any information at all, and information often had to be inferred by the readers, the reviewed articles rarely explain how and why teachers observed students - the field needs more rigorous approaches to study LS, studies on LS in initial teacher educations need to employ more coherent theoretical frameworks - frameworks such as noticing seem promising and should be further explored in the LS research - how teachers learn and observe needs to be better theorized and described in the research literature
Seleznyov	2018	systematic review	exploration of challenges to the implementation of LS beyond Japan	2005 - 2015	200	- there is a lack of a shared understanding of what LS entails outside of Japan - international LS adaptations frequently neglect or miss steps crucial to Japanese LS, especially those features that mark LS as a research process - LS adaptions are likely in new contexts, however, as LS is not structurally grounded in education systems outside of Japan - more research is needed into how and why LS is adapted in new contexts, and which consequences these adaptions may have

Seleznyov	2019	systematic review	evaluation of the LS impact outside of Japan (in-service teachers), and discussion of what type of impact measures are being used	2006 - 2016	56	- Findings concerning the evidence of impact (Guskey, 2000): Outcome 1: 57% reported that teachers were satisfied with LS Outcome 2: 80% of studies indicated positive changes in teachers' professional learning Outcome 3: 27% studies reported that LS had an impact on the school's organizational level, structures and professional learning culture Outcome 4: 59% of studies reported that LS had an impact on teachers' daily practice Outcome 5: 32% reported a positive impact on student learning - most of the reviewed studies are small-scale case studies lacking control groups and/or controlled variables, vast majority of the reviewed studies scored low on the Maryland Scientific Method Scale (MSMS) - the majority of studies were carried out in the US - there is weak evidence so far that LS has a lasting impact on teaches' daily practice, and their students' learning - studies that report on the impact of LS are predominately small-scale and have limited reliability due to their methodological design - the field needs large-scale evaluations of LS over time to collect robust evidence on the impact of LS
Willems & van den Bossche	2019	best evidence synthesis	evidence for the effectiveness of LS (focus on studies that used control groups)	2010 - 2018	5	- only 5 studies were identified that used well-controlled designs to investigate the impact of LS on teachers - more large-scale and long-term research is needed - the field needs to find better ways to measure teachers' professional learning - the field needs to identify circumstances and preconditions for learning
Kanello-poulou & Darra	2019	literature review	effectiveness of LS in higher education in Greece and internationally	2008 - 2018	28	- challenging areas in the implementation of LS with pre-service teachers are collaboration, lack knowledge, and inexperience with observation and reflection - crucial preconditions for the implementation of LS in higher education are preservice teachers' reflection and observation skills, their willingness to engage in collaboration, the support by teacher educators or mentors
Uştuk & Çomoğlu	2019	systematic review	LS in language teacher professional development in the Turkish context	2011 - 2018	11	- LS has the potential to support sustainable teacher professional development and teacher education in Turkey - they suggest to analyze existing LS research with the use of new theoretical frameworks and from the perspective of local realities
Rzejak	2019	systematic review	effectiveness of LS (in-service teachers)	2006 - 2016	9	- studies that provide evidence for the effectiveness of LS are scarce, the majority of LS publications are either narrations/reports or theoretical/ conceptual - no study concerning the effectiveness of LS in the Japanese context (and written in English) was eligible to be included in the review - the findings of eligible studies, despite indicating a positive impact of LS on teachers' knowledge, cannot be generalized
Norwich et al.	2021	systematic review	LS and LS-related practices in the field of special needs and inclusive education	all time	14 (LS) + 80 (related)	- the majority of studies used qualitative design, interviews, surveys or case study rather than experimental or controlled designs - the field should conduct more research into LS in inclusive education
Hervas	2021a	systematic review	history and spread of LS	up until 2006	unclear	- there are clear descriptions of LS in research papers also prior to Stigler & Hiebert 1999 - some of these studies are underrepresented in the literature
Hervas	2021b	systematic review	LS in higher education (faculty members)	1997 - 2019	21	- majority of the reviewed studies were conducted in the US - only few faculty members participated in most of the studies - there is an overall increase in studies on LS in higher education, yet overall, this stage of the education system is still underrepresented in overall LS research - several of the reviewed studies reported less optimistic findings about how the LS teams reflected and collaborated

Hrastinski	2021	integrative literature review	digital tools in LS	all time	13	- digital options can open up new ways to conceptualize and conduct LS - several tools have been tried out and can present diverse advantages - digital tools can help teachers prepare lessons and analyze observations, they can help to record LS and enable long-distance LS
Yoshida et al.	2021	systematic review	mapping the landscape of international LS research by classifying LS in themed groups and identifying trends	all time	891	- the majority of studies that aim to theorize LS or focus on the theoretical development of LS are carried out in Europe - different methodological approaches are used according to geographical region (e.g., predominant use of case study an open approach in Asia, action research used predominantly in US and Europe) - a high number of publications aim to describe or encapsulate misconceptions of Japanese LS, thereby upholding the idea that Japanese LS is the authentic and original way - LS is nowadays used in every stage of teacher education - Asia is the center of publications; however, US is usually the country with the most publications - the research focus in LS publications most commonly lies on mathematics, professional development, and teacher learning - there is a growing trend of studies with a focus on professional development that also contribute methodological suggestions
Gülhan	2021	systematic review	conceptualizations of LS in Turkey	all time	68	- most LS research in Turkey are with in-service teachers - majority of reviewed studies use qualitative methods and focus on mathematics
Wei & Huang	2022	literature review	LS and Learning Study in Asia (research-practice partnerships)	2011 - 2020	21	- there is a need to construct stronger interactions between research and practice, to identify shared problems and to reconceptualize the scripts of research and practice in Lesson Study and Learning Study
Baumfield et al.	2022	systematic review	how LS is used in initial teacher education	all time	51	- the majority of reviewed studies were carried out in the US - the majority of reviewed studies did not specify a methodological approach and some details about the LS intervention were missing (number of research lessons, review and planning meetings, learner focus, duration of LS practice, group members)
Fluminhan et al.	2022	systematic review	the relationship of LS, self-efficacy, and teacher learning	all time	8	- findings suggest that LS has several positive impacts on teachers' collaboration, their content knowledge, and self-efficacy - these benefits are limited by factors such as time pressure and lack of resources

Part II

Research Studies

Study 1

“We were thinking too much like adults”: Examining the development of teachers’ critical and collaborative reflection in Lesson Study discussions

Klara Kager, Anne Jurczok, Swantje Bolli, Miriam Vock

Kager, K., Jurczok, A., Bolli, S., & Vock, M. (2022). “We were thinking too much like adults”: Examining the development of teachers’ critical and collaborative reflection in lesson study discussions. *Teaching and Teacher Education*, *113*(103683), 1–13. https://doi.org/10.1016/j.tate.2022.103683

Abstract

This mixed-method study addresses the need for a clear conceptualization of the professional reflection element of Lesson Study (LS), a popular collaborative approach for the professional development of teachers. Grounding and re-framing LS's post-lesson discussion in a theoretical framework of critical and collaborative reflection, we analyze the transcripts of four LS groups at German primary schools, focusing on depth of reflection and teachers' trajectories through their reflective practice. The findings show that LS groups differed significantly in the depth and the trajectories of their reflection processes. We consider implications for post-lesson discussions and critical reflection as a LS core skill.

Keywords: teacher learning, professional development, critical and collaborative reflection, lesson study, critical inquiry

1. Introduction

Teachers are, without doubt, the primary agents of every educational system. They have to constantly stay on top of new pedagogies, teaching approaches, and educational reforms in order to equip students of diverging ability levels with increasingly complex competencies. This creates a need for professional development (PD) programs that provide in-service teachers with opportunities to continuously improve the quality of their professional practice. The iterative PD form Lesson Study (LS) is widely regarded as just such an effective learning model for teachers (Chokshi & Fernandez, 2004; Perry & Lewis, 2009). At the beginning of a LS cycle, teachers formulate a particular research interest and collaboratively plan a lesson or series of lessons that address this question. One member of the group then teaches the lesson while the others observe the students in the classroom. In the final step, the post-lesson discussion, teachers jointly describe, analyze, and discuss their observations, with the aim of arriving at future pedagogical objectives and translating the insights they have gained into improvements in their practice (Lewis, 2009).

LS is a highly collaborative PD approach, combining features of professional learning communities and collaborative teacher inquiry with the systematic and evidence-based analysis of jointly observed student learning. This collaborative analysis allows teachers to de- and reconstruct long-held assumptions about practice and to generate new knowledge through critical and collaborative reflection (Perry & Lewis, 2009). Importantly, the research lesson at the heart of the process is the product of the group's joint effort. It is not the respective teacher's skills that are scrutinized during the post-lesson discussion, but rather the students' responses to the lesson. These characteristics – collaboration, joint responsibility, and a critical lens that is not aimed at the teacher – have been identified as important preconditions for reflective conversations that may facilitate learning and promote change (e.g., Dewey, 1933; Hickson, 2011; Fook & Askeland, 2007; Nelson, et al., 2010).

Yet, while critical and collaborative reflection is an integral theoretical part of LS (Lewis & Tsuchida, 1998; Fernandez et al., 2003), it is far more challenging to achieve in practice. Evidence suggests that some groups use LS as a platform to reconfirm rather than question beliefs and practices (Wood, 2017), or struggle to transition from superficial reflection to critical reflection (Bae et al., 2016; Myers, 2012). The term *reflection,* despite being frequently used in research and practice, is marked by conceptual ambiguity and fuzzy boundaries between it and related educational concepts (DeLuca et al., 2015; Van Beveren et al., 2018).

To add to this, it is inherently difficult to demonstrate if and when reflection has in fact taken place, or to determine the quality of that reflection and its impact on knowledge building (Brown et al., 2021; Hatton & Smith, 1995). Similar challenges apply to research on professional change as well as teacher learning through reflection within LS (Mynott, 2019).

In order to address these issues, this collective case study employs a novel way of examining how different LS groups reflect together. The goal is to offer a theory-based definition of critical and collaborative reflection in the context of LS and apply this definition to the examination of reflective practices in post-lesson discussions. The study was conducted with in-service teachers in four German primary schools within the context of the "Leistung macht Schule"-project ("Excellence in School Education", BMBF & KMK, 2016). Using a mixed-method design, we ask the following research questions: (1) How do LS groups differ in the depth of their reflection in terms of reflective stages? (2) How do LS groups' respective trajectories through their reflective practice differ on a micro-diachronic scale?

2. Theoretical Framework

Within the field of education, critical reflection is predominantly conceptualized as a systematic and structured process of thinking that facilitates personal and professional growth (Fook & Askeland, 2007; Jordi, 2011). In contrast to reflection or introspection, critical reflection aims to not only understand the meaning of actions and situations, but the underlying assumptions and presuppositions governing one's actions and beliefs (Cranton, 1996; Mezirow, 1991). There is, however, no single definition of what this process entails, and terms such as *reflection*, *critical reflection*, and *dialogic reflection* are often used interchangeably (Hickson, 2011; Redmond, 2006). In fact, research has shown that the concept of reflection is frequently undertheorized in the social sciences and education (Brown et al., 2021; De Luca et al., 2015).

The undertheorization of reflection has not yet been established for the field of LS. There are, however, a number of indicators that suggest a lack of a shared understanding about the conceptualization and enactment of the reflection stage. Cerbin and Kopp (2006) note that there are no standards for how teachers collectively reflect and analyze their observations of student learning. This absence of standards is mirrored in the manifold approaches to structuring the reflection stage (e.g., Aji et al., 2018; Dudley, 2014; Knoblauch, 2019) and in

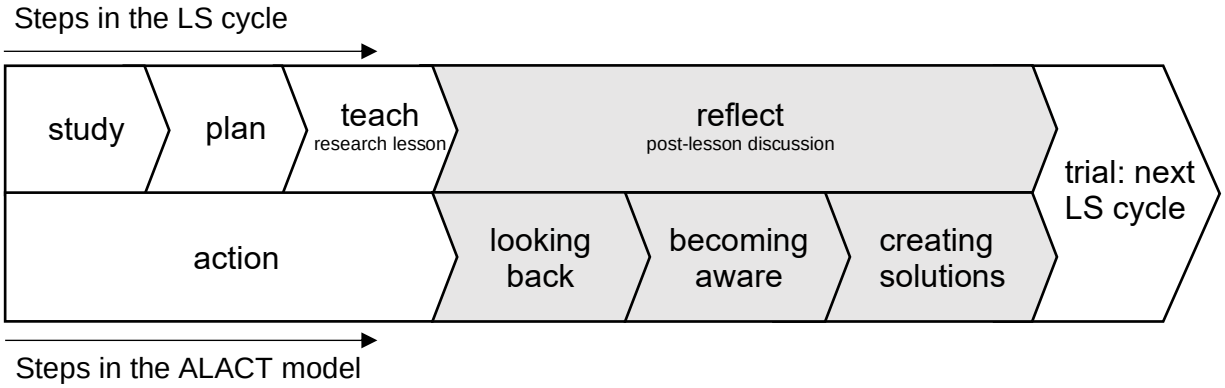
the diverse names allocated to it (e.g., “post-lesson discussion,” “colloquium,” “debrief meeting”).

There is also a diversity of theoretical perspectives on reflection in the research on post-lesson discussions in LS. Ricks (2011) and Bozkurt and Yetkin-Özdemir (2018), for example, adopt different approaches to exploring reflective activities. Ricks uses a process reflection framework based on Dewey (1981) and Schön (1983) and distinguishes between two reflective activities: reflection that is not connected to testing and reflection that involves the refinement of ideas through testing. Bozkurt and Yetkin-Özdemir (2017), on the other hand, summarize reflective activities under three headings: “evaluation,” “causal attribution,” and “inference.” Suratno and Iksandar (2010) and Myers (2012) both use Hatton and Smith’s (1995) levels of reflection, while Peña Trapero (2013) draws on the dimensions of practical thinking by Korthagen (2008). Alongside the great variety of conceptual approaches to reflection in LS, many articles provide minimal to no conceptualization on the reflection stage, nor how it was operationalized (e.g., Brosnan, 2014; Bruce & Hawes, 2015; Zhang et al., 2019).

Given the key role of reflection in teacher learning (Doğan & Adams, 2018) and a lack of consistency in its theoretical underpinnings and practical application in LS, we believe that it is important to identify a theoretical framework that could allow us to arrive at a shared definition and to operationalize reflection in real-life settings. As such, we suggest the ALACT model by Korthagen (1985) and Korthagen and Vasalos (2005), as it provides both a theoretical underpinning and simple practical guidance to re-think the post-lesson discussion. The ALACT model comprises five stages: (1) **a**ction; (2) **l**ooking back; (3) **a**wareness of essential aspects; (4) **c**reating alternative methods of action; and (5) **t**rial. These steps mirror a LS cycle (see Fig. 1). That is, teachers first plan and teach a research lesson (action), then come together to share their observations (looking back) and explain and analyze them (awareness of essential aspects). Finally, they try to formulate solutions (creating alternative methods). The insights from the post-lesson discussion are carried through into the next cycle in which teachers implement their pedagogical intentions or reframe their question (trial).

Figure 1

Mapping the Steps of a LS Cycle (Lewis, 2009) to the Steps of Korthagen's (1985) Stages of Reflection



The visualization in Fig. 1 demonstrates how the three middle stages of the ALACT model (looking back, becoming aware, and creating solutions) provide a script for how teachers can enact reflection in the post-lesson discussion. Further, we can now define critical reflection as consisting of three interconnected stages: looking back on and describing an event or experience; thereby becoming aware of one's own underlying assumptions and deconstructing their suitability and validity; and lastly deriving new insights and developing actions for the future (Korthagen, 1985).

The ALACT model conceptualizes how reflection can be enacted in theory. Research indicates, however, that the actual path through reflection takes various forms and does not always align with the systematic way proposed here. As a next step, we therefore want to consider some of these examples, beginning with findings from research on the post-lesson discussion.

2.1 The Post-Lesson Discussion: Affordances and Hinderances

The goal of the post-lesson discussion is for teachers to collaboratively generate new insights through an evidence-based analysis of student learning, and to arrive at alternative actions for an improved future practice (Dudley, 2013). Previous studies have identified several features that support the effectiveness of the post-lesson discussion. Clear guidelines and protocols can support the systematicity of the data collection and its subsequent analysis (Lewis et al., 2006; Lewis et al., 2019). Teachers benefit from going beyond play-by-play descriptions of how the lesson unfolded (Clevenger et al., 2009). It also helps to focus the

discussion on possible reasons for students' mistakes or misconceptions (Fernandez et al., 2003). Teachers' transition from play-by-play descriptions to a deeper analysis of observations can be supported by external facilitators through prompts and questions (e.g., Amador & Carter, 2018; Bae et al. 2016).

Dudley (2013) and Dudley et al. (2019) argue that LS groups can achieve most when "interthinking" and engaging in exploratory talk. Groups interthink when they use dialogue to create meaning as a collective and combine their cognitive resources to help them better explore problems (Mercer, 2001). Warwick et al. (2016) describe a pattern that is especially important to effective dialogue, and also closely aligns with the steps of the ALACT model (Korthagen, 1985). It occurs when teachers first describe an observation of individual student learning and then extrapolate this observation to a more general knowledge about learning. Thereby, teachers develop a shared understanding of the problem at hand and can work together towards suitable solutions. These steps are conceptualized as a "dialogic space" (Wegerif, 2007) for the development of pedagogical knowledge in LS and are seen as key to teacher learning (Warwick et al., 2016). The scarcity of these patterns in Warwick et al.'s (2016) data suggests, however, that teachers are often hindered when they try to effectively complete their journey through this dialogic space.

This is corroborated by studies that report that LS groups frequently converse on a superficial level or do not possess the skills to reflect critically (Brosnan, 2014; Callahan, 2019; Gutierez, 2015; Bae et al., 2016; Myers, 2012). LS groups might neglect crucial observations, change topics frequently, lead parallel discussions, and normalize problems as unchangeable situations that do not require further analysis (Karlsen & Helgevold, 2019). Other factors that may hinder a group's reflection process can be time pressure and disorganized record keeping (Lee & Tan, 2020), a predominant focus on teaching methods (Saito et al., 2006; Saito et al., 2008), and teachers' lack of sufficient professional knowledge (Lewis et al., 2009; Bae et al., 2016). These findings indicate that, while the goal of the post-lesson discussion may be straight-forward, much can go wrong when teachers attempt to critically reflect together as a group.

2.2 Reflecting Together as a Group – What Could Go Wrong?

In order to better understand the aspect of collaboration in teachers' reflections, we want to consider two frequently discussed assumptions related to reflection and apply them to the

collaborative reflection in LS. The first assumption is that reflection requires a problem, and the second relates to the focus of reflection.

Dewey (1933) and Korthagen and Vasalos (2005) argue that reflection requires a clear problem or discomfort, as well as the participants' willingness to experience this discomfort. What Dewey (1933) describes as discomfort may also be understood in terms of cognitive dissonance (Festinger, 1962; Piaget, 1975). Cognitive dissonance refers to a dissatisfaction when existing beliefs come into conflict with new experiences, which can motivate someone to examine their own belief system as well as identify and change deep-seated assumptions. Cognitive dissonance as a catalyst for conceptual change (i.e., the restructuring of mental concepts) is generally considered crucial to learning in both children and adults (e.g., Chan et al., 1997; Kang et al., 2004; Vosniadou et al., 2020).

There are several opportunities for teachers to individually and collectively experience cognitive dissonance over the course of a LS cycle. Teachers may face a dilemma in the classroom and thus decide to focus their research question on solving it. Dissonance may also emerge during the planning phase when teachers have diverging opinions on a topic, or during the research lessons when instructions in the classroom do not work as expected. The continuous exchange between teachers and external facilitators about their experiences can also be a powerful booster for dissonance and can stimulate discussions throughout all LS stages (Calleja & Formosa, 2020; Collet & Greiner, 2020). Research by Mynott (2019) has shown that it is also possible for LS groups to not experience dissonance at all, due, for instance, to a reluctance to engage in meaningful exchange, or to a general lack of interest in change. Calleja and Formosa (2020) stress that in order to develop moments of cognitive dissonance in LS, teachers need to feel safe to express their opinion and critically examine ideas within the group.

Research outside the field of LS lists several practices that can either support or constrain the collaborative critical examination of a problem. Slavit and Nelson (2010) describe conversations as consisting of several mini-inquiry cycles that involve "'doubling back' periods of readjustment" (p. 202–203). This means that sometimes groups have to reexamine proposed solutions and start over. Rodgers (2002) reports that some groups undergo "a series of intellectual dry runs through the problem/question and its various conclusions" (p. 854). In other words, disagreeing with each other in a constructive way can help a group to test out various paths in search of a viable solution.

Groups that tend to always agree with each other might jump to hasty conclusions or "quick-fixes," effectively omitting the stage of first identifying the underlying problem (Fook & Askeland, 2007; Korthagen & Vasalos, 2005; Schön, 1987). Horn and Little (2010) describe the phenomenon of normalizing, that is, when problems are treated as expected and as normal parts of practice. Normalizing a specific problem can help teacher groups to collectively focus their attention on the problem at hand and to critically consider general issues of teaching connected to this problem. However, the practice can also achieve the opposite effect when teacher groups limit their responses to reassurance and expressions of sympathy (Horn & Little, 2010). Together, these findings suggest that collaborative reflection is a cyclical process that requires participants to jointly identify a problem, actively build on each other and, if needed, reexamine solutions together in order to navigate the reflection stages of describing, explaining, and developing new ways forward.

The second assumption relates to the question posed by Korthagen and Vasalos (2005): "What does or should the teacher reflect upon?" (p. 51). Purposeful reflection needs a clear focus and objective, which should be co-constructed by all participants and based on a common vision (Nelson, 2009). This requires participants to work together and to risk exposing one's own shortcomings and misconceptions to oneself as well as to the group (Brookfield, 2017). This risk, also referred to as the "dark side" of reflection by Brookfield (2017), can lead to personal and emotional conflict between colleagues, or even damage a teacher's self-image. A resulting lack of trust between colleagues will likely lead to the stagnation of a group's inquiry, characterized by reconfirmation of beliefs already held, ego-protection, and disconnected talk (Brockbank & McGill, 2012; Nelson et al., 2010). These risks can be mitigated by trying to keep the object and content of reflection on an intellectual rather than personal level (Brookfield, 2017; Nelson et al., 2010).

The focus on student learning rather than the instructing teacher is an important hallmark of LS (Doig & Groves, 2011). This means that the post-lesson discussion concentrates on how students learned in the classroom. The LS group will likely observe more content than can realistically be explored in the limited time span of the post-lesson discussion and will have to jointly prioritize key observations. These observations should relate back to their research focus, otherwise teachers might simply "produce a laundry list" of unconnected classroom events (Lewis et al., 2019, p.31). It is also important that each group member perceives the focus of the post-lesson discussion as important to their own practice. Some members might

disengage from the discussion, if they feel that insights will not be transferable to their own classroom (Howell & Saye, 2016).

These examples illustrate that the process of critical and collaborative reflection will likely be impacted by a group's ability to identify problems and address them together in a focused inquiry. With this in mind, we use a mixed-method design to examine the post-lesson discussions of four LS groups according to (1) their depth of reflection in terms of reflective stages, and (2) their respective trajectories through their reflective practice.

3. The Present Study

This collective case study (Creswell et al., 2007) was conducted as part of a nationwide project organized by the German Research Association of the Federal Ministry of Education. Running from 2018 to 2023, the research alliance "Leistung macht Schule"- ("Excellence in School Education", BMBF & KMK, 2016). coordinates a variety of research projects across Germany with the aim of developing theory-based and field-tested concepts to support (potentially) high-achieving students. In our component project, we introduced LS to 19 primary schools, none of which had prior experience with the method.

Our PD concept combines on-site training with online learning materials, which follow the LS guidelines offered by Dudley (2014) and The Lesson Study Group at Mills College (**2018**). We adopted Knoblauch's (2019) learning activity curve as a method for observation and discussion. During the research lesson, teachers position themselves unobtrusively in the classroom and observe one or two case pupils each. They take detailed time-stamped notes using one sticky-note per observation (each child is assigned a different color). Case pupils are determined through high scores on pre-tests or are nominated by teachers. In the post-lesson discussion, the sticky-notes are arranged as coordinates on a wall or black board; the x-axis denotes the different phases of the lesson and the y-axis the learning activity of the case pupils. Teachers stick their notes higher or lower on the coordinate system, depending on whether the child appears involved or passive during a certain task (Knoblauch, 2019). This way, a child's learning over the different phases of the lesson can be visualized for all group members, and serve as evidence for the joint analysis. The process of arranging the sticky-notes further encourages teachers to connect their observation of a specific behavior or event with an interpretation of what this means for the child's learning.

The post-lesson discussion follows immediately after the research lesson and is usually conducted in an empty classroom. One of the external facilitators moderates the discussion and provides prompts to scaffold the reflection. The other facilitator acts as the record-keeper and notes down salient discussion points on white boards visible to all team members. Teachers are made aware of the intended phases of the discussion: the enacting teacher first shares their impression, the observing teachers then describe the learning activities of their case pupil(s) and hang the learning activity curves, all teachers collectively explain and analyze the observations, and finally teachers formulate consequences and solutions. To support this structure, we use three white boards for record keeping, labeled "Describing" (notes on descriptions of student learning), "Explaining" (notes on reasons and explanations for observations), and "Solutions" (notes on future actions). Both facilitators were present during each research lesson and acted as additional observers for two schools. All written notes and the learning-activity curve are photographed and digitalized for record-keeping for both the LS group and the facilitators. Teachers are encouraged to establish norms for their collaboration at the beginning of the LS process, but they do not receive any training in how to reflect together.

4. Methods

4.1 Data sample and collection

Using maximum variation selection strategy (Patton, 1990), we selected four primary schools (total sample = 19 schools) that differed in factors such as geographical location and in their experience with cooperative lesson planning and differentiation for (potentially) high-performing students. For the purposes of this study, we will refer to these four schools as school 1, 2, 3, and 4. All LS groups, consisting of four to eight teachers each (n = 24; 21 of them females), were conducting their first LS cycle. Data were collected in the first months of 2020. We created an audio-recording of the post-lesson discussion of each group (average length = 1.31 h; total audio data = 6.45 h). All participating LS group members gave informed consent for their audio-recordings to be used for research.

Each LS group chose their research question based on interest. Three groups focused on mathematics (grades 1, 2, and 4), and one on sciences (grade 6). For schools 2 and 3, the respective principals did not actively participate in the planning phases but joined the post-

lesson discussion to observe, listen, and sometimes share their opinion. The principals of schools 1 and 4 took more active roles by also observing a child in the research lesson. The principal of school 1 had also taken a leading role in the prior stages of the LS process and had planned the lesson together with the teachers.

Concerning the scope of the analysis, contextualization cues could not be taken into consideration. Disruptions, either internal (e.g., side conversations) or external (e.g., teacher leaving the room), were marked in the transcripts. Conversations about topics not relating to the research lesson (e.g., about logistics) were not coded. Table 1 presents a detailed overview of the sample.

Table 1

Overview of Data Sample

	school 1	**school 2**	**school 3**	**school 4**
discussion length (min.)	128	89	101	93
teachers	6	5	8	4
case pupils	4	4	5	4
grade	4th	1st	2nd	6th
subject	math	math	math	science
int. disruptions	16	78	5	15
ext. disruptions	6	5	3	3

Note. Int. disruptions refers to internal disruptions (e.g., teachers interrupting each other or engaging in parallel talk). Ext. disruption refers to external disruptions (e.g., teacher leaving the room).

The first and second author acted as external facilitators for all four LS processes. We first encountered LS in the context of this project and understand our role during the post-lesson discussion as moderators. We were not able to share expertise related to the content of the research lessons, but provided knowledge related to gifted education, differentiation, and LS.

4.2 Development of the Coding Tool

Some coding tools designed to capture teacher learning in LS have already been devised (e.g., Bae et al., 2016; Dudley, 2013; Vrikki et al., 2017; Warwick et al. 2016). Our initial plan

was to reuse the coding tool developed by Bae et al. (2016), as it captures the depth of teacher learning through hierarchical subcodes. We encountered several problems along the way, however, as we lacked access to detailed information on how the coding tool had been used. It also became apparent that, while existing coding tools, including Bae et al.'s (2016), do take the quality of teacher learning into consideration, none had been designed to explicitly examine the depth of teachers' reflection processes. In order to pursue our specific focus, we needed a coding tool that was also anchored in a theoretical perspective on reflection.

We therefore developed our own coding tool grounded in Korthagen's (1985, 2010) phases of reflection. The ALACT model was translated into hierarchical codes denoting the three phases of looking back (*Describing)*, becoming aware (*Explaining)*, and creating solutions (*Creating)*. Each code was inductively differentiated in two to four subcodes during the process of coding (see Table 2).

The coding process comprised several steps and followed qualitative content analysis (Schreier, 2012). First, the audio-files were transcribed verbatim using f4 (f4transkript v7 pro, 2020) and prepared for coding in MaxQda 2020 (Verbi Software, 2019). Each transcript was considered a unit of analysis, with the three phases of the post-lesson discussion treated as context units (Schreier, 2012). The coding unit was defined as a semantic unit, i.e., a unit of meaning.

In an initial joint coding session, two coders decided on the coding rules, derived the subcodes, and discussed how to code non-discriminatory segments and avoid double coding. Subsequently, the two coders worked independently and repeatedly compared and discussed their decisions. These comparisons led to various improvements of the coding tool, such as the deletion of ambiguous codes, refined definitions, and anchor examples. The code

Table 2

Coding Tool Based on the ALACT Model (Korthagen, 1985; Korthagen & Vasalos, 2005) Including Code Descriptions and Anchor Examples (T = teacher; P = principal)

Main code	Subcode	Description	Anchor example
1. Describing	**Describing_ Procedures**	Teachers describe procedures in the classroom (e.g., order or sequence of events, tasks, routines)	T: “At 12:07 she hands [her paper] in, goes back to her desk, doesn't clean it up, just relaxes for a moment, takes a sip and pours water all over herself <laughing> and then cleans herself up." (school 1, line 72)
	Describing_ Learning	Teachers describe how a child learns (learning processes, dynamics, habits, …)	T: “Concerning the next task she started with only one number, she started with 8, which you can see here. Then she suddenly realized, okay, what if I take 6 and 15, which makes 21. And she completely ignored the 8." (School 2, line 66)
2. Explaining	**Explaining_ Interpretation**	Teachers interpret or try to find meaning in a situation, an event, or behavior	T: “No, [the child] did look at you a bit, and, ehm, it seemed to me as if he wanted to make sure that everything was okay, I would say." (school 1, line 67)
	Explaining_ Explanation	Teachers elaborate on something or offer an explanation using facts or reasons not based on their own opinion	P: “He is a smart one. He made it into the top 20 in the regional math competition. So, he has a lot of potential.” (school 4, line 237)

	Explaining_ Interactive	Two or more teachers engage in a discussion, challenge or contradict each other, offer counterarguments (this code may span several utterances and is the only code that can be double coded with other subcodes)	T2: “We could include it in the consultation with the students.” T1: “I think that would be too much, we do not have enough time.” T3: “No, but we have a point on the agenda about what the students want to focus on in the future.” T4: “But then you have the responsibility of reading through all learning diaries before the consultation so that we know what to say to the students.” T1: “No, I think we should ... “ T3: “Yes, but in the learning diaries they write down 10 goals for their future work anyway [...]" (school 1, lines 723-724)
3. Creating	**Creating_ Realization**	Teachers summarize a discussion, formulate a conclusion or point out something new they have learned or realized	T: “That is the important thing, isn't it? To find a future path, that is the purpose of reflection." (school 1, line 765)
	Creating_ Solution	Teachers suggest a possible solution or a possible path forward	T: “I have an idea. We take the difficult task and let students start with that one, and then we provide them with an even more challenging one." (school 1, line 610)
	Creating_ Wish/ Intention	Teachers formulate a wish/ intention for their future practice	T: “It would have been really interesting to also observe a child with learning difficulties, to really see the contrast." (school 1, line 620)
	Creating_ DeeperQuestion	Teachers formulate a continuative or deeper question and/or anticipate future problems	T”: After we have used the learning diary regularly, we could say to the students in 4th grade: Okay, listen up, now you can formulate your own goals and come up with your own three indicators. Otherwise, we will never get to a point, where the students learn to set individual goals for themselves." (school 1, line 707)

Describing was relatively easy to recognize in the transcripts. Teachers focused on two main themes in their descriptions, which led to the development of the subcodes *Procedures* (e.g., description of tasks) and *Learning* (e.g., description of learning). The code *Explaining* proved far more difficult. The subcodes *Interpretation* (i.e., based on a teacher's opinion) and *Explanation* (i.e., based on facts or reason) were a first help to reliably recognize instances in the transcripts. There were several stretches, however, of messy discussion in the transcripts in which teachers collaboratively made sense of a situation. Within these stretches, teachers also used description and/or suggested solutions. In order to represent all codes, we created the code *Interactive*. This code is unique in the sense that it can span several utterances and may include several subcodes from various reflection stages. The last step, *Creating,* was enacted in various ways by teachers (e.g., concrete solutions, realizations, deeper questions, intentions). We thus develop four subcodes to represent this range.

Precise coding rules were set up in a coding manual after coding 25 percent of the data. After coding was completed, 30 percent of the transcripts (random sampling of passages from each transcript) were double-coded to calculate the percentage of agreement as a measure of coding consistency (O'Connor & Joffe, 2020). We used MaxQda to calculate the inter-coder reliability and achieved a reliability of 0.82 (Brennan's Kappa). All post-lesson discussions were conducted in German. The coding tool and excerpts discussed in this article have been translated to English.

4.3 Data Analysis

The data analysis combined a quantitative and qualitative approach. For the quantitative analysis, the raw frequencies of all codes were calculated in MaxQda and transformed into percentages. We counted instances and each of their possible repetitions, e.g., if a solution was mentioned several times throughout the conversation, each instance was counted separately. This means that if the overall frequency of the subcode *Creating_Solutions* is seven, the group did not necessarily produce seven distinct solutions, but talked about solutions seven times. This makes it possible to consider the actual frequency of certain reflective practices.

To compensate for the discussions' diverging lengths, we used percentages to represent the frequencies of all codes. The percentages were calculated by dividing the frequency of a given code by the overall frequency of all codes of the given school. Chi-square (χ^2) tests for

independence were used to compare the frequencies of codes between schools. Calculations were conducted in R (R Core Team, 2013), with the raw frequency of a given code as the observed frequency and the overall frequency of all codes for the given school minus the observed frequency as the expected frequency.

To analyze the schools' trajectories, we considered the post-lesson discussions holistically and on a micro-diachronic scale. In other words, we looked at the whole transcript instead of at selected episodes and analyzed how the discussion unfolded over time. This approach is informed by Pitzl (2020), who analyzed spoken interactions by developing participation profiles on a micro-diachronic scale. Specifically, the term micro-diachronic implies that the analytic focus lies on how a given conversation progresses in real time (Pitzl, 2020), such as over the course of a two-hour long post-lesson discussion. This means that even though the data sample is limited to a single conversation, we are focusing on the conversation's development on a microscale.

In order to visualize this development, we created a micro-diachronic portrait for each school. We used the code-line function in MaxQda for a sequential view of the coded segments. The frequencies of each main and subcode were calculated for every 30 speaker turns. This unit of analysis provides a balanced overview of all transcripts (exception: 60 speaker turns for school 1 due to length of discussion). In line with Slavit et al. (2013), we chose speaker turns rather than time as a unit of analysis and representative measure in order to minimize instances in which codes breach the unit boundary and would therefore have to be counted twice.

Concerning the qualitative analysis, we selected excerpts from the transcripts to complement the analysis (Kuckartz & Rädiger, 2019). Recognizing the danger of selective plausibility (Kuckartz, 2019), we selected examples that help to better understand the quantitative results. The excerpts are therefore not necessarily representative of the content of each post-lesson discussion.

5. Results

5.1 Quantitative Results

We briefly present the quantitative frequencies of the codes and highlight differences between their distribution within and across schools. We then describe the groups'

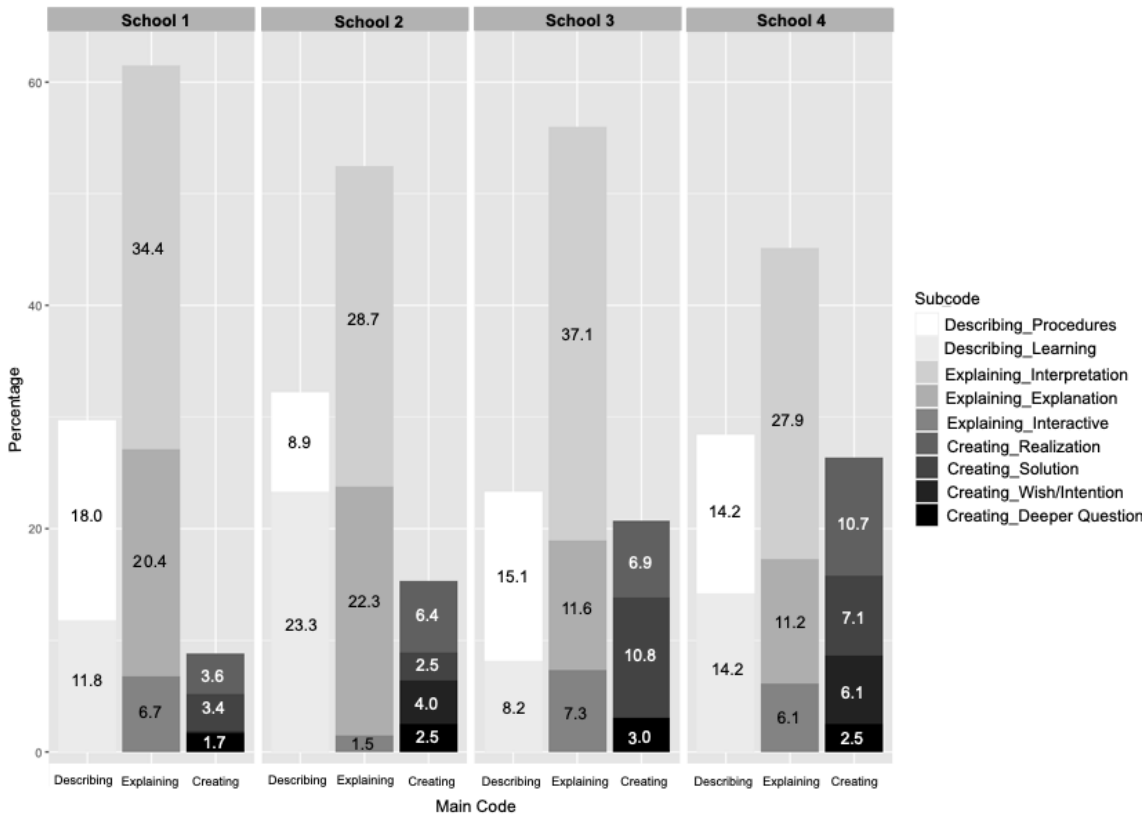
trajectories on a micro-diachronic scale. Concerning terminology, we speak of *stages of reflection* when referring to teachers' depth of reflection according to the ALACT model (*Describing*, *Explaining*, and *Creating*). These stages of reflection are independent from the *phases of the discussion*, which relate to the procedure and timing of the actual post-lesson discussion. The phases of the discussion (hanging the learning activity curves, explaining the curves, and creating solutions) were prompted by the facilitators.

5.1.1 Variance in Codes Between Schools

The frequency of codes provides descriptive evidence regarding how often groups either described events, explained them, or discussed future actions. Fig. 2 visualizes the percentages of subcodes across the three main codes and all four schools. Overall, nearly all subcodes are represented in each school, the exception being the absence of the code *Creating_Wish* for school 3. The graph shows that certain subcodes are more apparent than others in some schools.

Figure 2

Frequencies of Main and Subcodes in Percentages Across All Four Schools



Note. Values below 1 are not reported in the figure.

To support the descriptive results, a 2x4 χ^2 test for independence was conducted. The p-values were corrected for multiple comparisons using the Bonferroni-correction. Results indicate that there are statistically significant differences between schools for each main code (Table 3). To gain further insights into these differences, post-hoc analyses were conducted for each code.

Table 3

Main Code Frequency and Results for 2x4 χ^2 Tests of Independence Between Main Codes Across Schools

main code	school 1	school 2	school 3	school 4	chi-square test
Describing	30%	32%	23%	28%	χ^2 (3, N = 4) = 65.79, $p \leq .001$
Explaining	61%	52%	56%	45%	χ^2 (3, N = 4) = 16.94, $p \leq .001$
Creating	9%	15%	21%	26%	χ^2 (3, N = 4) = 41.52, $p \leq .001$

Describing. Fig. 2 indicates that school 2 had the highest percentage of the code *Describing*, and school 3 the lowest. A 2x2 post hoc analysis confirms that this difference between schools 2 and 3 is statistically significant (χ^2 (1) = 3.86, p = .049). Comparisons between the other schools did not yield any statistically significant results. Fig. 2 also shows a large variance in the distribution of the subcodes. Schools 1 and 3 focused on describing procedures, while school 2 predominantly focused on learning activities. School 4 shows an equal distribution between the subcodes.

Explaining. School 1 showed the highest percentage of the code *Explaining*, and school 4 the lowest; a difference approaching on significance (χ^2 (1) = 3.64, p = .056). Multiple 2x2 post-hoc analyses show that the differences between schools 1 and 2 (χ^2 (1) = 4.56, p = .032) and between schools 3 and 4 (χ^2 (1) = 4.60, p = .031) were significant. The remaining comparisons between schools were not statistically significant. Concerning the distribution of subcodes, school 2 was coded the least interactive; schools 1 and 3 the most interactive.

Creating. School 4 had the highest count of the code *Creating*, and school 1 the lowest. Multiple 2x2 post-hoc analyses indicated significant differences between schools 1 and 2 (χ^2

(1) = 5.99, p = .014), schools 1 and 3 (χ^2 (1) = 20.05, $p \leq .001$), schools 1 and 4 (χ^2 (1) = 36.69, $p \leq .001$), and schools 2 and 4 (χ^2 (1) = 6.73, p = .009). Concerning the subcodes, Fig. 2 indicates that school 4 had the highest percentage for the codes *Wish* and *Deeper Question,* while school 2 had the highest percentage of the code *Solutions*.

5.1.2 Micro-Diachronic Analysis: Variance in Development of Reflection Stages

The micro-diachronic portraits (Fig. 3) demonstrate that the established differences extend to the schools' trajectories through their reflective practice. We see that the three stages of reflection overlap with each other for each school. The schools differ in at least three aspects: the onsets of discussion phases, as well as the concurrence and overall balance of codes.

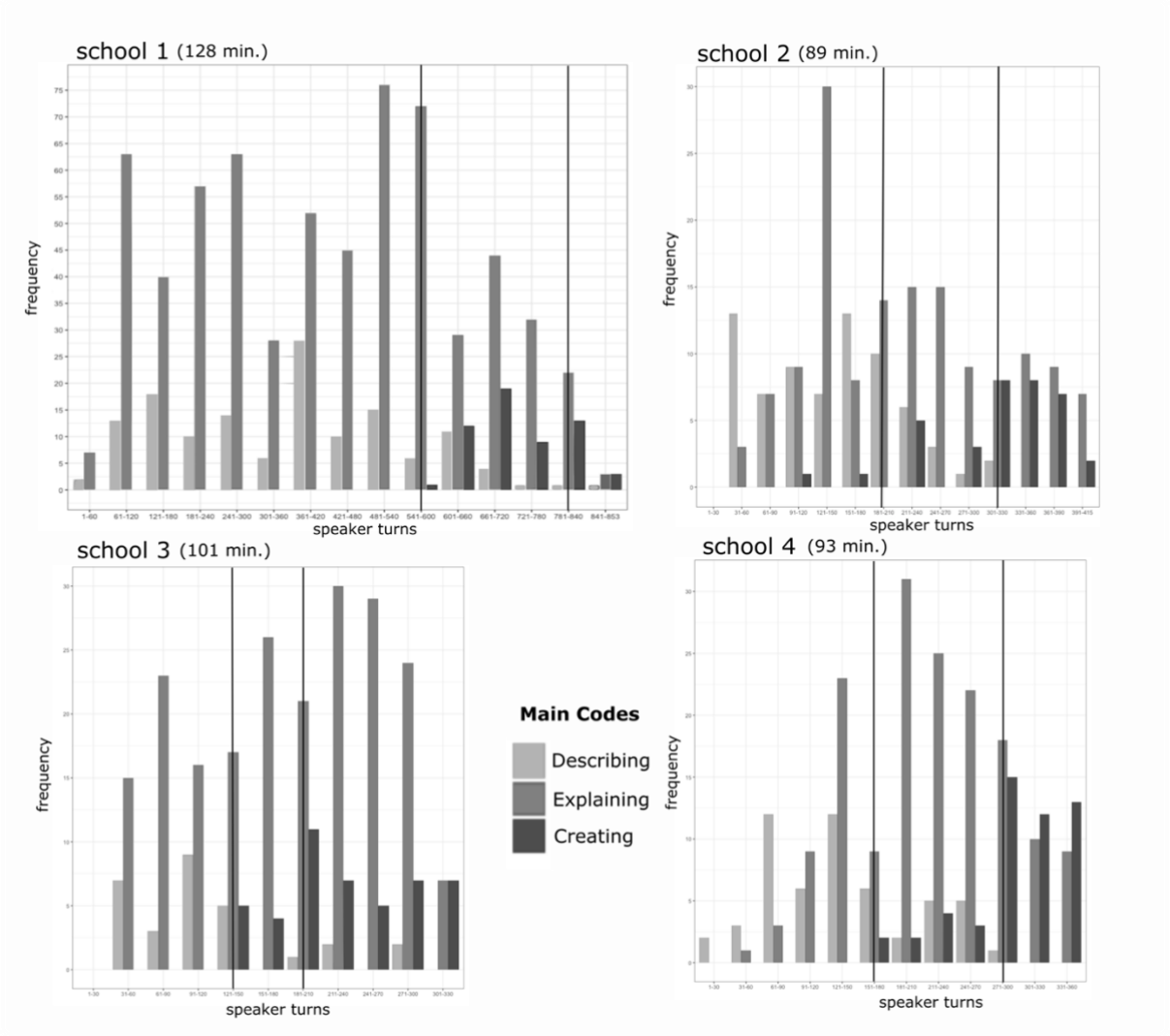
Concerning the first aspect – onset of discussion phases –, schools spent varying amounts of time on the first phase of their discussion. School 1 was the last to transition to the next phase of the conversation, whereas school 3 was the first. However, school 2 was the first to generate new ideas, even before transitioning to the second phase of the discussion. Schools also differed in the time they spent on the last phase of the conversation. School 3 transitioned first, spending approximately 42% of their discussion time talking about solutions. School 1, on the other hand, transitioned late, spending only approximately 7% of their conversation on this reflection stage. However, this portrait illustrates that school 1 was already creating ideas and solutions prior to the transition.

Turning to aspect two – the concurrence of codes –, schools 2 and 4 started their discussions on a generally descriptive level, with the code *Explaining* surpassing the code *Describing* at speaker turns 121–150 and 91–120 respectively. For schools 1 and 3 the code *Explaining* surpassed the code *Describing* from the onset of the conversation. The code *Describing*, highest in the beginning of each discussion, gradually decreased over time, but was still present during the second phase for each school.

Regarding the balance of codes – aspect three –, the illustration of school 2 demonstrates a peak of the code *Explaining* around speaker turns 121–150, indicating a vibrant discussion at this point during the conversation. The illustrations of the other schools appear more balanced overall.

Figure 3

Micro-Diachronic Portraits



Note. Distribution of main codes over the course of discussion for each school (30 speaker turn segments, exception: school 1, 60 speaker turn-segments). Vertical, black lines indicate prompts by facilitators to move to the next discussion phase (discussion phases: hanging of the learning activity curves, explaining and analyzing the learning activity curves, creating solutions for future practice).

5.2 Main Qualitative Results

The qualitative findings are structured along the phases of the post-lesson discussion and describe the LS groups' reflection processes during each phase. We first describe the phase of

hanging the learning activity curves, then the phase of explaining them, and finally the phase of creating future actions.

Phase 1: Hanging the Learning Activity Curves

The groups' first discussion phases varied in length and in the onset of the code *Explaining,* despite minimal differences in the number of case pupils. For schools 1 and 3, the code *Explaining* surpassed the code *Describing* from the word go. For schools 2 and 4, the first phase was clearly dominated by the code *Describing*, with *Explaining* emerging more gradually. A close analysis of the transcripts suggests that this difference might be due to the degree of interaction within the groups. For schools 1 and 3, this phase was characterized by teachers regularly probing clarifying questions or challenging their colleagues' understanding of learning activities. An example of the former is illustrated in the short exchange below (excerpt 1), in which the principal of school 1 describes and evaluates how a case pupil solved a task, and is asked for clarification by a colleague.

> *Excerpt 1.* School 1, lines 167-170 (P=principal, T=teacher)
>
> P: […] he writes down the solution by himself. He first discussed it with his partner, but then he writes it down independently. That is, in my opinion, a very independent, a very high learning activity.
> T4: I have a question. So, you say *independently*. With *independently* you mean that he discussed the solution with his partner but then they each wrote it down by themselves individually.
> P: Yes.
> T4: Instead of one child writing down [the solution] and the other one copying it.

This interactive character, however, might have contributed to school 1 lingering for too long on the discussion's first phase at the expense of the others. By the time the group was ready to explain the learning curve and create solutions, only limited time remained. Overall, this group showed an awareness of the reflection process by twice reminding themselves that they were "not supposed to interpret" yet (lines 133 and 513), and re-focusing their conversation back onto their research question (line 615).

By contrast, teachers of schools 2 and 4 generally reported their observations without any interruption or challenge from their colleagues, and also largely without asking for help

regarding the placement of the sticky-notes. These LS groups kept the first phase comparatively brief. A closer look at the transcript of school 2 hints, however, at a different problem: teachers failed to meaningfully connect their observations and interpretations and rarely built on each other's comments. Their discussion was characterized by 78 internal disruptions (in contrast to 16 for school 1), meaning that teachers continually interrupted each other or engaged in side-conversations. This is illustrated by the early peak in *Explaining*-codes (speaker turns 121–150), which represents the group's most interactive episode. A vigorous yet disconnected discussion emerged over the question of why some children need less scaffolding than others. Teachers, however, failed to take other opinions into consideration, continuously interrupted each other and finally dropped the topic.

Phase 2: Explaining the Learning Activity Curves

For all schools, the second phase of the discussion was marked by instances of all three reflection stages. Notably, the code *Describing* remained present throughout this phase; and even beyond for schools 1 and 3. Teachers frequently repeated descriptions or added previously unreported details relevant to a new line of discussion, thereby reinforcing their argument or reminding the group of the topic's initial starting point.

The code *Creating* emerged immediately for all schools when they transitioned into this phase, and for school 2 even earlier. If we look at excerpt 2 below, we can see that the teachers in school 3 suggested solutions directly after the facilitator prompts them to explain their observations:

> *Excerpt 2.* School 3, lines 139-142 (T=teacher)
>
> T1: In my opinion the overall learning activity was really high, especially during the partner work. Ehm, even if they didn't reach the solution that we wanted them to reach […]. Ehm, maybe the instructions should have been more specific. Or [the pupils] should have received advice-cards earlier on, or there should have been the advice from our side that they should look at them and use the material from the beginning on […] maybe then there would have been more activity and they might have reached the solution we wanted them to reach.
>
> [two more teachers offer unconnected explanations]
>
> T4: Wait, let's go back again […]

In the excerpt, a group member identifies the problem of the research lesson (pupils did not manage to solve the task) and instantly offers solutions, answered by two other teachers' unrelated suggestions. Finally, another teacher suggests that they start over. The appeal to return to the original question helped the group to focus the conversation back on the problem. Following a short exchange on how the case pupils approached the task, the LS group suddenly reached a very different understanding of what had transpired in the classroom (excerpt 3).

Excerpt 3. School 3, lines 148 and 169, (T=teacher)

> T5: Thinking about it, I don't think it was a problem that the pupils didn't solve the task and that the task description was … a bit open and that it didn't prescribe exactly how they should reach the solution, or what solution they should reach.
>
> […]
>
> T5: […] for some [pupils] it was actually good, that they really had to work [on the problem], and not take the easier route of just getting an advice-card.

The reexamination of the problem helped teachers to no longer perceive the fact that none of the pupils had reached the expected solution as a failure, but as a valuable learning opportunity. By the end of this phase, the group had pinpointed several factors that had challenged the pupils' learning, such as the limited time frame and the pupils' lack of familiarity with open tasks and the provided materials. One teacher summarized the groups' approach to planning the lesson with the words "we were probably thinking too much like adults" (line 257). With this in mind, the group subsequently agreed on several actions to better equip pupils with the skills necessary to independently solve challenging tasks and work through the related frustration.

Phase 3: Creating Future Actions

Concerning the last phase, in which teachers aim to arrive at a shared understanding and agree on next moves, we want to focus on two instances from school 2. These instances illustrate the need for the critical examination of observations in order for teachers to also create viable solutions. Already early on in their discussion, two teachers expressed their

surprise over the fact that most pupils displayed high learning activities despite the research lesson taking place during fifth period, which is late in the school day. When another teacher described that a case pupil had worked at a slower pace than usual, the group jointly attributed this to the lesson taking place during fifth period. The group's general belief that pupils' learning activities decrease over the course of the day prevented any further exploration of possible reasons for this decrease in activity. The pupils' slow working pace was treated as an inevitable fact that did not need solving and could be sufficiently explained by the fifth-period-argument.

The second instance concerned another case pupil that had scored highly on a pre-test and received a weekly plan for advanced exercises. During the research lesson, the child accidently took the exercise sheet for an easier level and worked on the wrong exercises for most of the lesson. When discussing the incident, the teachers had several explanations ready, such as "this is typical for this child," "the child is always absent-minded/ doesn't listen/ is inattentive and overeager," and "it's a language problem," (school 2, summary of lines 239–260). The "typical" and "always" suggest that the teachers had noticed similar situations before, but had normalized the child's problems as an unchangeable behavior, meaning that the child's struggle was "old news". The teachers did not problematize the child's need for further scaffolding and soon changed the subject, despite having identified relevant explanations (such as a language barrier).

The group agreed on some future actions by the end of the discussion. These actions were largely based on reconfirmations of beliefs rather than on evidence-based arguments. The group was nevertheless of the general opinion that they had learned a lot during their LS cycle. Given the inherent difficulty of pinpointing learning, we cannot infer whether or not school 2's post-lesson discussion led to professional learning. We can, however, pinpoint these two instances of potential learning that the group neglected or did not possess the joint skills to cultivate.

6. Discussion

This paper addressed two research questions, both predicated on the finding that some LS groups struggle to engage in critical and collaborative reflection (Bae et al., 2016; Myers, 2012). Specifically, we asked (1) how do LS groups differ in the depth of their reflection in

terms of reflective stages?, and (2) how do LS groups' respective trajectories through their reflective practice differ on a micro-diachronic scale?

The analysis found instances of the three stages of reflection (Korthagen, 1985) in each discussion, indicating that each LS group went beyond play-by-play descriptions of events. The micro-diachronic portraits also showed that each discussion generally followed the stage's hierarchical sequence. The transitions between stages, however, were hardly discernible. Even with the same two facilitators providing prompts for transitions, the discussions developed along dynamic and unique trajectories. This is consistent with the view that stages of reflection are hard to distinguish from one another, and that transitions are often seamless (Dewey, 1944; Rodgers, 2002). A closer look into the quantitative frequencies of codes also revealed statistically significant differences in the distribution of codes across schools. This means that, while all schools reached the stage of transforming observations into actions, they did so in different ways.

Differences in the Schools' Reflection Processes

The results (Fig. 2) indicated that the LS groups differed in what they described. Teachers of school 2 concentrated their descriptions on how students had learned (23.3%) over classroom procedures (8.9%). Previous literature suggests that post-lesson discussions benefit from a clear focus on student learning over other issues, such as teaching methods (Doig & Groves, 2011; Saito et al., 2006). We saw, however, that teachers of this school struggled to develop their detailed descriptions of student learning. Teachers added more and more experiences and observations to the discussion, instead of connecting them to a line of argument. In the end, important observations, such as the one of the girl that had mistakenly worked on the wrong exercises, remained mere observations.

Teachers of school 3, on the other hand, concentrated their descriptions on classroom procedures (15.1%) over student learning (8.2%). The LS group nevertheless managed to pinpoint salient observations of student thinking and successfully connected them back to their research theme. The group led a focused and critical reflection about the students' struggle and why the LS group had not anticipated this struggle. We suggest that our results emphasize Lewis et al.'s (2019) argument that quality observations and their recounting in the post-lesson discussion do not guarantee a quality discussion. Instead, LS groups need the skills to collectively exploit their observations in order to reflect critically.

We also saw differences in how LS groups approached the task of hanging the learning activity curves. Schools 1 and 3 simultaneously and collaboratively described and interpreted classroom observations. Schools 2 and 4 mainly described observations and rarely discussed where to place them on the curve. The value of following the subsequent steps of reflection has been emphasized repeatedly across the literature, as it can improve the systematicity of the inquiry (e.g., Hatton & Smith, 1995; Van Es & Sherin, 2002). Keeping the teachers' task during the hanging of the learning activity curves in mind, however, we believe that the interactive and interpretative stance visible for schools 1 and 3 function as an advantage rather than a disadvantage in their conversations. The method of the learning activity curve asks teachers to not only describe, but to evaluate a pupil's cognitive learning activity (Knoblauch, 2019). Voicing and questioning opinions in this phase facilitates the groups' collective inquiry into aspects of their own beliefs about learning and allows the development of a shared vocabulary.

Another difference between schools' trajectories concerns the frequency of the code *Creating* (see Fig. 2). Teachers of school 1 spent only 8.7% of their discussion on this stage of reflection, whereas teachers of school 4 allocated 20.4% to it. These numbers invite the conclusion that teachers of school 4 underwent a deeper critical reflection than teachers of school 1. The qualitative analysis of all schools, however, showed that issues might be discussed superficially and proposed solutions will not necessarily be the result of deep reflection. For example, we observed teachers jumping to quick fixes, and collectively normalizing situations by limiting their responses to standard solutions and avoiding critical inquiry (Horn & Little, 2010; Fook & Askeland, 2007). Teachers of school 1 might have agreed on a lower number of future moves compared to the other schools, yet the teachers had co-constructed, evaluated, and critically examined the viability of each of them.

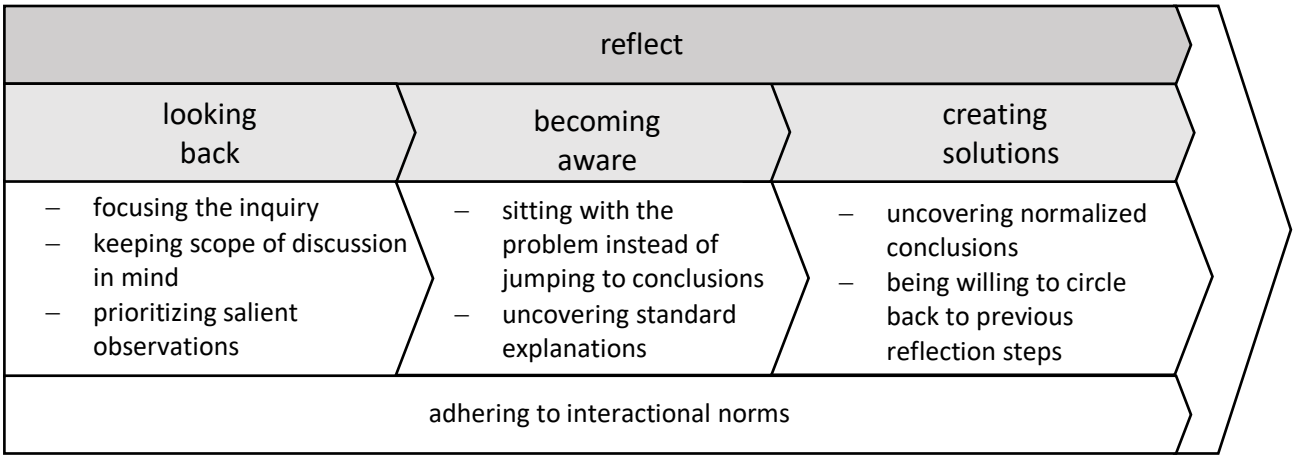
This means that considering only the overall frequencies of a certain code might skew our perception of a LS group's success in engaging in critical reflection. Our results suggest that there is more to reflection than simply following a protocol, such as the ALACT model. The reflection stages of describing, explaining, and creating helped teachers of schools 1 and 3 to self-regulate their inquiry and circle back to analyze the root causes of some of the problems. The protocol did not have the same effect on schools 2 and 4, even though all three reflection stages were present in their discussions as well. These findings indicate that following the ALACT model can support groups in their reflection processes. Concerning the analysis of

groups' reflection processes, however, the presence of the models' stages in the post-lesson discussion was not a reliable indicator of the depth of discussion.

The analysis presented in this paper provides valuable insights into how four LS groups at German primary schools differed in their critical and collaborative reflection and in their respective trajectories on a micro-diachronic scale. The specific difficulties that shaped the LS groups' trajectories are summarized in Fig. 4.

Figure 4

Summary of Difficulties that Shaped the Trajectories of Schools



Implications for the Post-Lesson Discussion

As a next step, we want to discuss what the findings summarized in Fig. 4 mean for the effective progression of the post-lesson discussion. The analysis indicates, in our eyes, that a LS groups' critical and collaborative reflection depends on at least three interdependent aspects: a transparent structure or protocol, the facilitators' awareness of different trajectories through reflection, and the teachers' skills to reflect critically in a group.

Firstly, the post-lesson discussion should be based on a transparent structure that aids the development of the reflection process by including stages such as those lined out in the ALACT model. A well-structured post-lesson discussion can help to keep the stages of reflection intact and remind teachers not to jump to conclusions too quickly or resort to default explanations. We provided minimal instructions to our teachers concerning the stages of reflection, yet the key words "Describing," "Explaining," and "Solutions" written on whiteboards helped at least

two groups to systematize their inquiry even without facilitator intervention (schools 1 and 3). This suggests that already simple protocols can help promote awareness of structured inquiry.

Secondly, external facilitators need to be aware of the various trajectories reflection in the post-lesson discussion can take, and of the challenges that LS groups may face. Our findings indicate that some LS groups require sustained help in initially establishing and enforcing the norms of interaction. Even LS groups that appear adept to reflecting critically in a group might still need assistance concerning time-management, focusing their discussion, and keeping its scope. Facilitators might also need to decelerate a group's reflection process to prevent hasty conclusions. This task could be supported by the explicit structure or protocol mentioned above, which the facilitator can refer to.

Thirdly, teachers need to possess the skills necessary to navigate the steps of critical and collaborative reflection. These skills include the ability to act and think as a group, to tolerate and productively develop discomfort, and to systematically question actions and situations – even familiar ones (e.g., Korthagen & Vasalos, 2005; Mynott, 2019; Nelson, 2009). Evidence has shown that LS groups should not be expected to bring all these skills to the table fully formed (e.g., Lewis et al., 2006; Myers, 2012). Our analysis confirms these findings and demonstrates that protocols alone cannot make up for inexperience with critical and collaborative reflection. We therefore argue that the advancement of reflection skills has to be an integral part of LS. There is a considerable amount of literature on the skills that teachers develop *through* LS, but little discussion on the skills teachers should possess *before* entering into LS. Critical reflection as a group can be a powerful tool for teachers to interthink and lead evidence-based discussions, yet its procedures and mechanisms need to be acquired and routinized (Korthagen & Vasalos, 2005).

Reflection skills become especially critical when considering how schools can continue self-sustained LS beyond the duration of a project. In order for teachers to lead and guide LS without external expertise, they need to take charge of their own professional development and cultivate a set of core skills that will support their learning. We should therefore include the *before* into our consideration of teacher learning in the LS process. Which skills do teachers need to possess, in order to be able to capitalize on learning opportunities in LS? And, how can these skills be bolstered in the LS context? To take the discussion by Mynott (2019) concerning observation and feedback as LS core skills further, we argue that investing

into the advancement of teachers' reflection skills is likely to maximize the opportunity for professional learning in LS and may promote the sustained translation of teacher learning to pupil learning.

Limitations

Although this study found several meaningful results, it is subject to limitations. Firstly, the analysis is based on only four schools. A bigger sample might have led to even clearer differences in the schools' trajectories and exposed further difficulties that shaped these patterns. Given our mainly qualitative approach, we can only speculate why the groups' trajectories through their reflection process differed and whether these differences have an impact on the group's learning. Future studies should look more closely at the conditions and effects that influence a LS group's reflection process.

Our coding tool was designed to count discourse units – an approach that can be criticized for oversimplifying human interaction (Slavit et al., 2013). The micro-diachronic analysis was plotted along speaker turns rather than actual time. While this methodological practice is believed to produce more representative measures (Slavit et al., 2013), the portraits presented in this paper should still be interpreted with caution.

Lastly, our analysis is tightly linked to the method of the learning activity curve (Knoblauch, 2019). Even though we deem this approach helpful for teachers' critical and collaborative reflection, it is not to say that other approaches to structuring the post-lesson discussion cannot also provide this guidance in equal or stronger form.

7. Conclusion

In this paper, we put forward a theory-based definition of critical and collaborative reflection in the context of LS and provided a novel way of visualizing the reflection process over the course of post-lesson discussions. Our analysis indicated that post-lesson discussions follow dynamic and unique routes. It confirmed that there is no single way to reflect and delineated several challenges to the reflection process. Given the goal of LS to transform observations into practicable knowledge, we argue that rethinking the post-lesson discussion in terms of reflective stages, such as those proposed in the ALACT model, can make this goal more attainable for teachers, as well as sustainable for long-term teacher-led LS. This was illustrated by the example of a LS group that uncovered their preexisting assumption that

planning a lesson “thinking too much as adults” would meet the learning needs of the pupils. However, the findings demonstrated that protocols alone cannot make up for a lack of reflection skills. That is, the ability to jointly deconstruct already held beliefs and collectively pursue ideas in a focused discussion that ventures beyond standard explanations and solutions.

Future research should aim to develop a more comprehensive picture of how LS groups around the globe engage in critical reflection, and how this process can be best scaffolded through structures, procedures, and strong LS core skills. For example, we should consider existing concepts or programs for facilitating critical reflection and investigate their usefulness in the context of LS.

References

Aji, S. D., Hudha, M. N., Huda, C., Nandiyanto, A. B. D., & Abdullah, A. G. (2018). The improvement of learning effectiveness in the lesson study by using e-rubric. *Journal of Engineering Science and Technology*, *13*(5), 1181–1189.

Amador, J. M., & Carter, I. S. (2018). Audible conversational affordances and constraints of verbalizing professional noticing during prospective teacher lesson study. *Journal of Mathematics Teacher Education*, *21*(1), 5-34. https://doi.org/10.1007/s10857-016-9347-x

Bae, C. L., Hayes, K. N., Seitz, J., O'Connor, D., & DiStefano, R. (2016). A coding tool for examining the substance of teacher professional learning and change with example cases from middle school science lesson study. *Teaching and Teacher Education*, *60*, 164-178. https://10.1016/j.tate.2016.08.016

BMBF, & KMK. (2016). *Gemeinsame Initiative von Bund und Ländern zur Förderung leistungsstarker und potenziell besonders leistungsfähiger Schülerinnen und Schüler*. [Joint initiative by the federation and federal states for the support of gifted and potentially gifted pupils]. https://www.bmbf.de/de/leistung-macht-schule-3641.html

Bozkurt, E., & Yetkin-Özdemir, İ. E. (2018). Middle School Mathematics Teachers' Reflection Activities in the Context of Lesson Study. *International Journal of Instruction*, *11*(1), 379–394. https://doi.org/10.12973/iji.2018.11126a

Brockbank, A., & McGill, I. (2012). *Facilitating reflective learning: Coaching, mentoring and supervision* (2nd ed.). Kogan Page Publishers.

Brookfield, S. (2017). *Becoming a critically reflective teacher.* John Wiley & Sons.

Brosnan, A. (2014). Introducing lesson study in promoting a new mathematics curriculum in Irish post-primary schools. *International Journal for Lesson and Learning Studies*, *3*(3), 236–251. https://doi.org/10.1108/IJLLS-09-2013-0050

Brown, C., Poortman, C., Gray, H., Groß Ophoff, J., & Warf, M. (2021). Facilitating collaborative reflective inquiry amongst teachers: What do we currently know? *International Journal of Educational Research*, *105*, 101695. https://doi.org/10.1016/j.ijer.2020.101695

Bruce, C. D., & Hawes, Z. (2015). The role of 2D and 3D mental rotation in mathematics for young children: What is it? Why does it matter? And what can we do about it? *ZDM – Mathematics Education*, *47*(3), 331–343. https://doi.org/10.1007/s11858-014-0637-4

Callahan, C. (2019). Middle school geography teachers' professional development centered around historical photographs. *The Journal of Social Studies Research*, *43*(4), 375–388. https://doi.org/10.1016/j.jssr.2018.11.003

Calleja, J., & Formosa, L. (2020). Teacher change through cognitive conflicts: The case of an art lesson study. *International Journal for Lesson & Learning Studies*, *9*(4), 383–395. https://doi.org/10.1108/IJLLS-05-2020-0028

Chan, C., Burtis, J., & Bereiter, C. (1997). Knowledge building as a mediator of conflict in conceptual change. *Cognition and Instruction, 15*(1), 1-40. https://doi.org/10.1207/s1532690xci1501_1

Chokshi, S., & Fernandez, C. (2004). Challenges to importing Japanese lesson study: Concerns, misconceptions, and nuances. *Phi Delta Kappan*, *85*(7), 520-525. https://doi.org/10.1177/003172170408500710

Cerbin, W., & Kopp, B. (2006). Lesson Study as a Model for Building Pedagogical Knowledge and Improving Teaching. *International Journal of Teaching and Learning in Higher Education*, *18*(3), 250–257. https://www.isetl.org/ijtlhe/pdf/ijtlhe110.pdf

Clevenger, M., Kuhnley, S., O'Rourke, C., & Umland, K. (2009). Visualizing interaction during lesson study: A graph-theoretic approach. *Journal of Classroom Interaction, 44*(1), 39-47. http://www.jstor.org/stable/23869289

Collet, V. S., & Greiner, A. C. (2020). Revisioning Grammar Instruction through Collaborative Lesson Study: A New Apprenticeship of Observation. *Literacy Research and Instruction*, *59*(2), 95–120. https://doi.org/10.1080/19388071.2019.1709927

Cranton, P. (1996). *Professional development as transformative learning: New perspectives for teachers of adults.* Jossey-Bass.

Creswell, J. W., Hanson, W. E., Clark Plano, V. L., & Morales, A. (2007). Qualitative Research Designs: Selection and Implementation. *The Counseling Psychologist*, *35*(2), 236–264. https://doi.org/10.1177/0011000006287390

DeLuca, C., Shulha, J., Luhanga, U., Shulha, L. M., Christou, T. M., & Klinger, D. A. (2015). Collaborative inquiry as a professional learning structure for educators: A scoping review. *Professional Development in Education*, *41*(4), 640-670. https://doi.org/10.1080/19415257.2014.933120

Dewey, J. (1933). *How we think: A restatement of the relation of reflective thinking to the educative process*. D.C. Heath & Co.

Dewey, J. (1944). Some questions about value. *The Journal of Philosophy*, *41*(17), 449-455. https://doi.org/10.2307/2020017

Dewey, J. (1981). How we think. In J. A. Boydston (Ed.), *John Dewey: The later works, 1925–1953* (Vol. 8, pp. 105–352). Southern Illinois University Press.

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), 634–659. https://doi.org/10.1080/09243453.2018.1500921

Doig, B., & Groves, S. (2011). Japanese lesson study: Teacher professional development through communities of inquiry. *Mathematics Teacher Education and Development*, *13*(1), 77-93. https://files.eric.ed.gov/fulltext/EJ960950.pdf

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*, *34*, 107-121. https://doi.org/10.1016/j.tate.2013.04.006

Dudley, P. (2014). *Lesson Study: A handbook for teachers and school leaders*. Lesson Study UK. Retrieved from www.lessonstudy.co.uk (First published 2003 by NCSL).

Dudley, P., Xu, H., Vermunt, J. D., & Lang, J. (2019). Empirical evidence of the impact of lesson study on students' achievement, teachers' professional learning and on institutional and system evolution. *European Journal of Education*, *54*(2), 202-217. https://doi.org/10.1111/ejed.12337

F4transkript. (2020). *Audiotranskription F4transkript* (v7 pro). Dr. Dresing & Pehl GmbH – audiotranskiption. https://www.audiotranskription.de/en/f4transkript/

Fernandez, C., Cannon, J., & Chokshi, S. (2003). A US-Japan lesson study collaboration reveals critical lenses for examining practice. *Teaching and Teacher Education, 19*(2), 171-185. https://doi.org/10.1016/S0742-051X(02)00102-6

Festinger, L. (1962). *A theory of cognitive dissonance* (Vol. 2). Stanford University Press.

Fook, J., & Askeland, G. A. (2007). Challenges of critical reflection: 'Nothing ventured, nothing gained'. *Social Work Education*, *26*(5), 520-533. https://doi.org/10.1080/02615470601118662

Gutierez, S. B. (2015). Teachers' reflective practice in lesson study: A tool for improving instructional practice. *Alberta Journal of Educational Research*, *61*(3), 314-328. https://doi.org/10.11575/ajer.v61i3.56087

Hatton, N., & Smith, D. (1995). Reflection in teacher education: Towards definition and implementation. *Teaching and Teacher Education, 11*(1), 33–49. https://doi.org/10.1016/0742-051X(94)00012-U

Hickson, H. (2011). Critical reflection: Reflecting on learning to be reflective. *Reflective Practice*, *12*(6), 829-839. https://doi.org/10.1080/14623943.2011.616687

Horn, I.S., & 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), 181-217. https://doi.org/10.3102/0002831209345158

Howell, J. B., & Saye, J. W. (2016). Using lesson study to develop a shared professional teaching knowledge culture among 4th grade social studies teachers. *The Journal of Social Studies Research*, *40*(1), 25–37. https://doi.org/10.1016/j.jssr.2015.03.001

Jordi, R. (2011). Reframing the concept of reflection: Consciousness, experiential learning, and reflective learning practices. *Adult Education Quarterly*, *61*(2), 181-197. https://doi.org/10.1177/0741713610380439

Kang, S., Scharmann, L. C., & Noh, T. (2004). Reexamining the Role of Cognitive Conflict in Science Concept Learning. *Research in Science Education*, *34*(1), 71–96. https://doi.org/10.1023/B:RISE.0000021001.77568.b3

Karlsen, A. M., & Helgevold, N. (2019). Lesson study: Analytical stance and depth of noticing in post-lesson discussions. *International Journal of Lesson and Learning Studies, 8*(4), 290-304. https://doi.org/10.1108/IJLLS-04-2019-0034

Knoblauch, R. (2019). Beobachtung und Dokumentation von Lernaktivitäten in der Lesson Study [Observation and documentation of learning activity in lesson study]. In C. Mewald & E. Rauscher (Eds.), *Lesson Study. Das Handbuch für kollaborative Unterrichtsentwicklung und Lernforschung* [*Lesson study. The handbook for collaborative lesson development and learning research*] (pp. 49-76). Studienverlag.

Korthagen, F. A. (1985). Reflective teaching and preservice teacher education in the Netherlands. *Journal of Teacher Education*, *36*(5), 11-15. https://doi.org/10.1177/002248718503600502

Korthagen, F. A. (2008). *Linking Theory and Practice. The Pedagogy of Realistic Teacher Education*. Routledge.

Korthagen, F. A. (2010). Teacher reflection: What it is and what it does. In E.G. Pultorak (Ed.), *The purposes, practices, and professionalism of teacher reflectivity: Insights for twenty-first-century teachers and students* (pp. 377-401). Rowman & Littlefield Education.

Korthagen, F. A., & Vasalos, A. (2005). Levels in reflection: Core reflection as a means to enhance professional growth. *Teachers and Teaching: Theory and Practice, 11*(1), 47-71. https://doi.org/10.1080/1354060042000337093

Kuckartz, U. (2019). Qualitative Text Analysis: A Systematic Approach. In Kaiser, G., & Presmeg, N. (Eds.). (2019). *Compendium for Early Career Researchers in Mathematics Education* (pp. 181-198). Springer International Publishing. https://doi.org/10.1007/978-3-030-15636-7

Kuckartz, U., & Rädiker, S. (2019). *Analyzing Qualitative Data with MAXQDA: Text, Audio, and Video*. Springer International Publishing. https://doi.org/10.1007/978-3-030-15671-8

Lee, L. H. J., & Tan, S. C. (2020). Teacher learning in lesson study: Affordances, disturbances, contradictions, and implications. *Teaching and Teacher Education*, *89*, 1-15. https://doi.org/10.1016/j.tate.2019.102986

Lewis, C., & Tsuchida, I. (1998). A lesson is like a swiftly flowing river: Research lessons and the improvement of Japanese education. *American Educator, 22*(4), 12-17, 50-52. https://tinyurl.com/nr98pvdp

Lewis, C. (2009). What is the nature of knowledge development in lesson study?. *Educational Action Research*, *17*(1), 95-110. https://doi.org/10.1080/09650790802667477

Lewis, C., Friedkin, S., Emerson, K., Henn, L., & Goldsmith, L. (2019). How Does Lesson Study Work? Toward a Theory of Lesson Study Process and Impact. In R. Huang, A. Takahashi, & J. P. da Ponte (Eds.), *Theory and Practice of Lesson Study in Mathematics* (pp. 13–37). Springer International Publishing. https://doi.org/10.1007/978-3-030-04031-4_2

Lewis, C., Perry, R., Hurd, J., & O'Connell, M. P. (2006). Lesson study comes of age in North America. *Phi delta kappa*, *88*(4), 273-281. https://doi.org/10.1177/003172170608800406

Lewis, C. C., Perry, R. R., & Hurd, J. (2009). Improving mathematics instruction through lesson study: A theoretical model and North American case. *Journal of Mathematics Teacher Education*, *12*(4), 285-304. https://doi.org/10.1007/s10857-009-9102-7

Mercer, N. (2001). Why I study interthinking. *The Psychologist, 14*(12), 648-649. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.629.6369&rep=rep1&type=pdf

Mezirow, J. (1991). *Fostering critical reflection in adulthood: A guide to transformative and emancipatory learning.* Jossey-Bass.

Myers, J. (2012). Lesson study as a means for facilitating preservice teacher reflectivity. *International Journal for the Scholarship of Teaching and Learning*, *6*(1), 1-21. https://doi.org/10.20429/ijsotl.2012.060115

Mynott, J. P. (2019). Lesson study outcomes: A theoretical model. *International Journal for Lesson and Learning Studies, 8*(2), 117-134. https://doi.org/10.1108/IJLLS-08-2018-0057

Nelson, T.H. (2009). Teachers' collaborative inquiry and professional growth: Should we be optimistic?. *Science Education, 93*(3), 540-580. https://doi.org/10.1002/sce.20302

Nelson, T. H., Deuel, A., Slavit, D., & Kennedy, A. (2010). Leading deep conversations in collaborative inquiry groups. *The Clearing House*, *83*(5), 175-179. https://doi.org/10.1080/00098650903505498.

O'Connor, C., & Joffe, H. (2020). Intercoder Reliability in Qualitative Research: Debates and Practical Guidelines. *International Journal of Qualitative Methods*, *19*, 160940691989922. https://doi.org/10.1177/1609406919899220

Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). Sage Publications, Inc.

Peña Trapero, N. (2013). Lesson study and practical thinking: A case study in Spain. *International Journal for Lesson and Learning Studies*, *2*(2), 115–136. https://doi.org/10.1108/20468251311323379

Perry, R. R. & Lewis, C. C. (2009). What is a successful adaptation of lesson study in the US?. *Journal of Educational Change, 10*(4), 365-391. https://doi.org/10.1007/s10833-008-9069-7

Piaget, J. (1975). *L'e´quilibration des structures cognitives. Proble`me central du de´velopment [The development of thought: equilibration of cognitive structures].* Presses universitaires de France & Viking Press.

Pitzl, M. L. (2020). Tracing the emergence of situational multilingual practices in a BELF meeting: Micro-diachronic analysis and implications of corpus design 1. In K. Murata

(Ed.), *ELF research methods and approaches to data and analyses: theoretical and methodological underpinnings* (pp. 97-125). Routledge.

R Core Team. (2013). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. http://www.R-project.org/

Redmond, B. (2006). *Reflection in action: Developing reflective practice in health and social services* (2nd ed.). Ashgate.

Ricks, T. E. (2011). Process Reflection during Japanese Lesson Study Experiences by Prospective Secondary Mathematics Teachers. *Journal of Mathematics Teacher Education*, *14*(4), 251–267. https://doi.org/10.1007/s10857-010-9155-7

Rodgers, C. (2002). Defining reflection: Another look at John Dewey and reflective thinking. *Teachers College Record, 104*(4), 842–866. https://tinyurl.com/yzn52snk

Saito, E., Harun, I., Kuboki, I., & Tachibana, H. (2006). Indonesian lesson study in practice: case study of Indonesian mathematics and science teacher education project. *Journal of in-service education, 32*(2), 171–184. https://doi.org/10.1080/13674580600650872

Saito, E., Tsukui, A., & Tanaka, Y. (2008). Problems on professional development of primary teachers in Vietnam: Case of Bac Gian province. *International Journal of Educational Development*, *28*(1), 89–103. https://doi.org/10.1016/j.ijedudev.2007.08.001.

Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. Basic Books.

Schön, D. A. (1987). *Educating the reflective practitioner: Toward a new design for teaching and learning in the professions*. Jossey-Bass.

Schreier, M. (2012). *Qualitative content analysis in practice*. Sage Publications, Inc.

Slavit, D., & Nelson, T. H. (2010). Collaborative teacher inquiry as a tool for building theory on the development and use of rich mathematical tasks. *Journal of Mathematics Teacher Education*, *13*(3), 201-221. https://doii.org/10.1007/s10857-009-9136-x

Slavit, D., Nelson, T. H., & Deuel, A. (2013). Teacher groups' conceptions and uses of student-learning data. *Journal of Teacher Education*, *64*(1), 8-21. https://doi.org/10.1177/0022487112445517

Suratno, T., & Iskandar, S. (2010). Teacher reflection in Indonesia: Lessons learnt from a lesson study program. *US-China Education Review*, *7*(12), 39–48. https://files.eric.ed.gov/fulltext/ED514886.pdf

The Lesson Study Group at Mills College. (2018). *Conduct a Cycle*. https://lessonresearch.net/conduct-a-cycle/overview/

Van Beveren, L., Roets, G., Buysse, A., & Rutten, K. (2018). We all reflect, but why? A systematic review of the purposes of reflection in higher education in social and behavioral sciences. *Educational Research Review*, *24*, 1-9. https://doi.org/10.1016/j.edurev.2018.01.002

Van Es, E. A., & Sherin, M. G. (2002). Learning to notice: Scaffolding new teachers' interpretations of classroom interactions. *Journal of Technology and Teacher Education*, *10*(4), 571-596. https://tinyurl.com/d736nx8y

VERBI Software. (2019). *MAXQDA 2020*. VERBI Software. https://www.maxqda.com

Vrikki, M., Warwick, P., Vermunt, J. D., Mercer, N., & Van Halem, N. (2017). Teacher learning in the context of lesson study: A video-based analysis of teacher discussions. *Teaching and Teacher Education*, *61*, 211-224. https://doi.org/10.1016/j.tate.2016.10.014

Vosniadou, S., Lawson, M. J., Wyra, M., Van Deur, P., Jeffries, D., & Ngurah, D. I. G. (2020). Pre-service teachers' beliefs about learning and teaching and about the self-regulation of learning: A conceptual change perspective. *International Journal of Educational Research*, *99*, 101495. https://doi.org/10.1016/j.ijer.2019.101495

Warwick, P., Vrikki, M., Vermunt, J. D., Mercer, N., & 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*, *48*(4), 555-569. https://doi.org/10.1007/s11858-015-0750-z

Wegerif, R. (2007). *Dialogic education and technology: Expanding the space of learning* (Vol. 7). Springer Science & Business Media.

Wood, K. (2017). Is there really any difference between lesson and learning study? Both focus on neriage. *International Journal for Lesson and Learning Studies, 6*(2), 118-123. https://doi.org/10.1108/IJLLS-02-2017-0008

Zhang, H., Yuan, R., & Liao, W. (2019). Teacher Development Facilitated by Lesson Study: A Chinese Perspective. *TESOL Quarterly*, *53*(2), 542–552. https://doi.org/10.1002/tesq.480

Study 2

A Systematic Review of Transparency in Lesson Study Research: How Do We Report on the Observation and Reflection Stages?

Klara Kager, Eva Kalinowski, Anne Jurczok & Miriam Vock

This study has been submitted for publication and is currently under review.

Kager, K., Kalinowski, E., Jurczok, A., & Vock, M. (submitted*). A Systematic Review of Transparency in Lesson Study Research: How Do We Report on the Observation and Reflection Stages.*

Abstract

Lesson Study is a method of professional development for teachers that has gained traction in the past few decades. The method is, however, undertheorized and research indicates that publications routinely fail to describe crucial details of the intervention's implementation. This challenges the meaningful synthesis and replicability of research findings. Using a protocol based on Moravcsik's (2020) dimensions of transparency, this systematic review examines 129 articles on Lesson Study published between 2015 and 2020 to identify how transparent their reporting of how teachers observed and reflected together was. The large-scale findings confirm a lack of transparency across several dimensions of how the Lesson Study intervention is reported and highlight a current lack of theorization and coherence in the field. To address some of these issues, we propose a framing structure that empirical papers on Lesson Study should give critical attention to in order to ensure relevance and transferability.

Keywords: professional development, lesson study, research transparency, observation, reflection, systematic review

Introduction

Lesson Study (LS) is a popular approach to the collaborative professional development (PD) of teachers. Originating in Japan, LS has seen a surge of international interest in the past three decades (Yoshida et al., 2021). LS consists of iterative cycles in which a group of teachers follows a series of core stages: study, plan, teach and observe, and reflect (Lewis, 2009). Specifically, teachers identify a problem or question relevant to their practice, they then study the curriculum about that issue and subsequently plan a lesson or series of lessons that address it. Following this phase, one teacher teaches the lesson, while the remaining group members observe the lesson with a focus on student learning. These observations are subsequently analyzed, with teachers collaboratively reflecting on their lesson and negotiating alternative approaches for future teaching (Lewis, 2009). Numerous case studies (e.g., Coenders & Verhoef, 2019; Schipper et al., 2018), but also some randomized controlled trials (e.g., Lewis & Perry, 2015, 2017), indicate that LS can represent a powerful way for teachers to engage in PD and enhance various types of knowledge. A rising number of studies, however, also report mixed or less promising results (e.g., Bjuland & Mosvold, 2015; Canonigo, 2016; Parks, 2008), suggesting that LS is not always successful and can result in a variety of different qualitative outcomes.

This discrepancy in the benefits of this PD approach raises the question of the degree to which the way that LS is implemented influences the impact it has on teachers. Several case studies on LS report that teachers sometimes struggle to systematize their observations and reflect critically (e.g., Bae et al., 2016; Bjuland & Mosvold, 2015; Karlsen & Helgevold, 2019), while others suggest that the use of templates and protocols could scaffold these stages of the LS intervention (e.g., Færøyvik Karlsen, 2019; Kager et al., 2022). Despite being understood as a systematic and evidence-based approach to improving teaching, LS actually has no such standardized methods or unified protocols (Cerbin & Kopp, 2006). Instead, the literature frequently speaks of key features and core stages (e.g., Seleznyov, 2018) that can be adapted by researchers and practitioners to fit their specific needs and cultural contexts. LS research thus reports on highly contextualized and diverse versions of the intervention (Hadfield & Jopling, 2016), which makes reaching a conclusive synthesis of evidence on the efficacy of the method very difficult.

The challenges outlined above emphasize the need for a shared language as well as high transparency in scientific articles concerning the way in which the given LS intervention was

implemented. Findings by Larssen et al. (2018), however, indicate that LS literature on initial teacher education tends to omit information on how teachers observed student learning in LS, and frequently lacks a clear theoretical foundation for the measure. Their findings hint at a bigger issue that goes beyond the field of LS: the importance of clearly communicating steps and decisions taken during an intervention, and more broadly during research. Several review articles have shown that the research literature in the social sciences, including the field of education, frequently falls short of this cornerstone of transparent communication (Aguinis et al., 2018; Brown et al., 2021; DeLuca et al., 2015; Hardwicke et al., 2020; Mann & Walsh, 2013).

The starting point of this paper is the position that a clear theoretical foundation and thorough and transparent descriptions of LS interventions in publications is necessary to move the field of LS forward. Based on findings by Larssen et al. (2018) concerning LS literature on initial teacher education, we hypothesize that this theorization and transparency is currently lacking in literature on LS with in-service teachers. The present review therefore sets out to address the way we conceptualize and communicate LS interventions in empirical research. We use Moravcsik's (2020) three dimensions of research transparency to conceptualize how decisions taken by researchers and teachers can be communicated transparently. By means of a systematic review, we examine how the observation and reflection stages of LS are currently reported in peer-reviewed in-service teacher literature over the past five years. Specifically, we aim to answer the following questions: (1) How transparent are LS articles in reporting their observation and reflection stages? And (2) which theoretical frameworks are used in these studies to conceptualize the observation and reflection stages in LS? Based on our findings, we then propose a framework for how prospective empirical articles can best report on the observation and reflection stage in LS.

The Need for Rich Description in Lesson Study

The field of LS has grown exponentially in the past few decades (Yoshida et al., 2021), but several reviews of the literature demonstrate that there remains a lack of robust evidence for LS's effectiveness (Cheung & Wong, 2014; Seleznyov, 2019; Willems & Van den Bossche, 2019). Both Cheung and Wong (2014) and Seleznyov (2019) note that only a few articles on the efficacy of LS use a well-controlled experimental design. Large-scale experimental or quasi-experimental studies arguably provide "the most valid and scientifically defensible evidence" (Guskey & Yoon, 2009, p.498) for the effectiveness of a PD. The few examples of

such studies in the field of LS, however, illustrate Bryk's (2015) argument that experimental studies can be challenging to design, time-intensive, and expensive. In addition, the type of knowledge they produce may not always be sufficient in practice, as implementation fidelity of a PD in a different context cannot always be guaranteed (Bryk, 2015; Lewis, 2015).

This is why other type of research on PDs, such as qualitative approaches, are also a valuable source of learning in educational science. This perspective, often referred to as improvement science (Bryk, 2015), suggests that in the absence of standardization we can utilize variability to better understand which factors of a PD might lead to which impact and outcome in which contexts (Bryk, 2015). If we want to systematically learn from variation and synthesize reliable evidence for LS's effectiveness through qualitative research, however, we need theory-based definitions and rigorous, comparable, and replicable descriptions of the LS interventions. The review by Larssen et al. (2018) found that these rigorous descriptions are currently lacking in publications on LS in initial teacher education. Examining 34 articles, the authors found that the majority of studies made little or incoherent use of theoretical frameworks and frequently treated details about the LS implementations with a "taken-for-granted understanding" that required readers to "infer how the observation had been conducted" (Larssen et al., 2018, p. 17).

While Cheung and Wong's (2014) and Seleznyov's (2019) reviewes addressed the design and research methods of LS articles, Larssen et al. (2018) examined how the LS intervention was reported. Despite their similar objectives, these studies thus focused on different aspects that are both crucial parts of LS research. These are, first, the study's scientific research methods (i.e., case study, randomized controlled trial), and, second, the LS intervention itself (i.e., how teachers executed the core stages of LS intervention). Especially in studies on interventions, such as LS, both aspects need to be thoroughly described in order to assess both the quality of the research, as well as the quality of the intervention itself. Without clear descriptions of the specific intervention and how it was executed, readers can neither compare results to those of other studies using the intervention, nor replicate the intervention in their own context (Rosenshine, 1994).

Our review of the literature indicates that the LS field currently struggles with both routes to reliable evidence for the PD's efficacy: large-scale experimental evidence is rare, while the growing body of small-scale qualitative research seems to lack transparency concerning how the LS intervention was conceptualized by researchers and executed by teachers. We can only

start to discern LS outcomes and the likely factors that contributed to them, when we situate research within a shared theoretical framework and use mutually intelligible terminology and descriptions. This means that LS research needs to be reported in a transparent way, which includes replicable and comaprative descriptions of the LS itervention.

A Framework for Transparency in Lesson Study Research

The issue of transparency in the field of LS research has received little attention to date, yet the topic is gaining traction in light of the replication crisis in psychology (Wigginss & Christophersosn, 2019) and as part of the discourse around Open Science (Makel & Plucker, 2014; van Dijk et al., 2021). Research transparency refers to "the degree of detail and disclosure about specific steps, decisions, and judgement calls made during a scientific study" (Aguinis et al., 2018, p. 84). In other words, studies display high research transparency if they explicitly communicate choices made by the researchers about design, data collection, and analysis, and if they make resources, such as protocols and materials, available. Especially in qualitative research, there seems to exist some confusion over how research transparency can be best achieved for different types of research (Moravcsik, 2020), which has resulted in "a serious neglect of transparency and reproducibility" in some parts of social sciences, including education (Hardwicke et al., 2020, p. 7). Wiggins and Christopherson (2019), approaching transparency from the angle of the replication crisis that has hit psychology in the last decade, note that the way in which data is collected and analyzed cannot be treated as a "secret recipe" (p. 209), but has to be replicable to others. Studies that report on an intervention, such as LS, therefore need to include detailed descriptions of the intervention's design and execution in order for other researchers to be able to transfer the intervention.

There are several reasons for why critical information concerning a study's methodology or intervention might get lost on the journey to publication. First, the omission of information might be due to external circumstances, that is, some information may be subject to ethical or legal barriers, or has to omitted due to the strict word limits that some journals have (Moravcsik, 2020). Second, researchers might expect their readers to understand certain terms or processes without further explanation. Frequently used constructs are often presumed to be understood universally, at least among researchers in a specific discipline (Eisenhart & DeHaan, 2005; Wolgemuth et al., 2017). As a result, these constructs tend to be underdefined in the literature, often lacking a theoretical underpinning. One such construct

"riddled with inconsistencies" (Mann & Walsh, 2013, p. 292) in the field of education is "reflection" (e.g., Brown et al., 2021; DeLuca et al., 2015; Mann & Walsh, 2013). While articles generally identify reflection as being a vital part of teachers' inquiry processes, the reflection process itself remains largely undefined in publications (DeLuca et al., 2015) and descriptions of *how* reflective practice can be operationalized are routinely omitted (Mann & Walsh, 2013). These findings are reiterated in the recent meta-narrative literature review by Brown et al. (2021) on reflective professional inquiry, which shows that the undertheorization of reflection, while increasingly criticized, still very much exists in the field of education. The use of the term 'reflection' in a research article without a definition or theoretical grounding is therefore not particularly helpful to readers and challenges the works' transparency and replicability.

Guidelines by journals or, for instance, the *Guide to APA Style* (American Psychological Association, 2022), provide clear recommendations on how to report method sections in scientific papers. In addition, research has produced lists and recommendations for how transparency can be improved in different fields and in specific types of research papers (e.g., Hardwicke et al., 2020; Meyrick, 2006). Moravcsik (2020), focusing on social sciences, delineates three normative dimensions of research transparency that can help us to better conceptualize which aspects contribute to a clear description of research. The first dimension, *data transparency*, concerns access to data and evidence that researchers base their findings on. Access to data enables other researchers to fully understand the analysis at hand and to judge its validity, as well as to improve or extend that analysis (Moravcsik, 2020). The second dimension, *analytic transparency*, concerns the way in which data has been collected and analyzed. This dimension is especially critical in qualitative research, as "social scientific evidence does not speak for itself" (Moravcsik, 2020, p. 3), but has to be inferred. The third dimension, *production transparency*, concerns the wider contextual conditions that impacted the collection and analysis of data—in other words, the methodological choices and processes that led to these choices.

These three dimensions have been formulated for the assessment of the transparency of research methods. As noted early, several studies have already focused on research methods and design in LS research (Cheung & Wong, 2014; Seleznyov, 2019; Willems & Van den Bossche, 2019). Larssen et al.'s (2018) review of how the LS intervention is reported in initial teacher education indicates that further research into the transparency of the LS intervention

is needed. We therefore narrow down our analysis to the degree of transparency by which the stages of observation and reflection of the LS intervention are communicated in research. In a next step, we discuss how Moravcsik's (2020) framework of research transparency, which was formulated to assess research methods, can help us better understand which aspects of the LS intervention are crucial to report in the literature.

Which Aspects of Lesson Study Need to be Communicated?

During the stages of observation and reflection, teachers conduct systematic observations on student learning and then analyze them collectively. While this process resembles a research process—and despite teachers often being encouraged to adopt a research stance in their LS work (Lewis et al., 2019)—, teachers' observations and reflections in LS are not subject to scientific standards. The descriptions of these processes in research publications, on the other hand, are a crucial part of the intervention and need to be transparently communicated. This is especially important since several studies describe that the way in which the stages of observation and reflection were executed and supported by materials can have an effect on what teachers learn from the experience and whether they perceive LS as useful to their work (Bae et al., 2016; Bjuland & Mosvold, 2015; Karlsen & Helgevold, 2019). By using Moravcsik's dimensions (2020), we can identify and systematize steps in teachers' observation and reflection processes that might influence the interventions' outcome and thus need to be transparently recorded in publications.

Translated to the LS process, the dimension of *data transparency* concerns the observation stage, in which teachers observe and record student learning. Both Brosnan (2014) and Bjuland and Mosvold (2015) describe cases in which the overall quality of the LS cycle suffered in part from teachers' unstructured note-taking. Færøyvik Karlsen (2019) and Callahan (2019), on the other hand, describe that the use of specific observation protocols enhanced teachers' observations. In order for other researchers to reconstruct teachers' observation process, articles therefore need to be clear on *how* (e.g., unstructured notes, specific template) and *whom* (e.g., whole class, case pupils) teachers observed, as well as about the materials that scaffolded this process (e.g., lesson plan, video recordings, phones).

The second dimension, *analytic transparency*, concerns the reflection stage in LS, in which teachers aim to derive new approaches for their future practice based on their observations (Lewis et al., 2019). Several studies describe that LS groups found it challenging to reflect

critically together (e.g., Bae et al., 2016; Myers, 2012), or create and develop potential moments of learning in their discussions (Mynott, 2019). LS literature should therefore explicitly report back on how the teachers approached the reflection stage and whether their collaborative reflection followed a specific structure or protocol.

The third dimension, *production transparency*, includese a broader set of aspects that may influence the observation and reflection stage: outside expertise, the way in which LS groups document their LS process, and the duration and setting of the reflection stage. The involvement of external expertise, in the form of, say, knowledgeable others and external facilitators, is an integral part of LS and has often been shown to play a crucial role in how impactful the measure is on teachers (e.g., Amador & Carter, 2018; Bae et al., 2016). The extent of their involvement in the LS process is therefore an important factor that needs to be described in research studies. Furthermore, LS is not a one-time event, but relies on iteration (Stigler & Hiebert, 2016). In order for LS groups to be able to consolidate their learning, transfer it to their next LS cycle, or be able to communicate their findings to their school and wider community, it is vital that they keep some kind of a record of their learning (Lewis et al., 2019; Seleznyov et al., 2021). In more recent studies, time and space for teachers' reflection have been highlighted as being important preconditions for succesful LS (Seleznyov et al., 2021).

The Present Study

The present study aims to systematically review empirical research on LS with in-service teachers. It has two objectives. Firstly, we aim to verify whether the findings by Larssen et al. (2018) concerning the underreporting of the observation stage in LS literature on initial teacher education holds true for literature on in-service LS, which represents the bulk of LS literature. We extend Larssen et al.'s (2018) focus on the observation stage to the reflection stage, as reflection has been shown to be frequently undertheorized und underdescribed in research studies (Brown et al., 2021; DeLuca et al., 2015). Secondly, we aim to synthesize whether the stages of observation and reflection are connected to, or defined in relation to, a theoretical framework. We ask the following questions:

(1) How transparent are LS articles in reporting their observation and reflection stages of LS?

(2) Which theoretical frameworks are used in these studies to conceptualize the observation and reflection stages in LS?

By giving critical attention to the issue of transparency and theorization in LS articles, we hope to instigate an open dialogue on the issue in the LS community and beyond. We aim to set this dialogue in motion by recommending a framework based on our findings on the reporting of the observation and reflection stages in empirical LS articles.

Method

We followed the stages of a systematic review as set out by Gough (2007, pp. 218–19) and have structured this section accordingly. We first define the inclusion criteria, then delineate the search strategy, and finally describe the coding process and data analysis. Prior to data analysis, we developed a systematic review protocol based on the PRISMA checklist proposed by Moher et al. (2009). The review protocol, along with a version of the coding tool, were pre-registered on Open Science Framework (OSF) on November 22, 2021, and both are available at https://doi.org/10.17605/OSF.IO/5NXGY (Kager et al., 2021).

Inclusion Criteria

The review included an article if: (a) it reported on LS with in-service teachers in a general educational school or preschool (kindergarten to secondary school); (b) it was published in a peer-reviewed journal; (c) it was published between January 2015 and December 2020; (d) it was available in English; (e) it was an original and predominantly qualitative study; (f) it focused on LS (rather than on a PD approach that only includes elements of LS); and (g) it focused on either the whole LS process or specifically on the observation and reflection stages.

The review protocol published on OSF provides a detailed account of our rationale behind each of these inclusion criteria. Nevertheless, we want to highlight and explain some decisions we made during the culling process. To begin with, we initially focused on studies published within the last decade (2010 to 2020) and in doing so identified an overwhelming number of eligible studies (see Figure 1). To keep the body of studies to a manageable size, and given that the majority of identified studies had been published between 2015 and 2020, we shortened the time frame to this period.

Secondly, we focused on qualitative research as it represents the bulk of LS research (Xu & Pedder, 2014) and one would expect this kind of research to be most likely to include detailed

descriptions of the LS interventions in question. We therefore included several qualitative designs, such as narrative research, case study, grounded theory, phenomenology, participatory action research, design-based research, and action research. Large-scale implementations of LS were therefore excluded, but have previously been reviewed in studies with a similar focus (Seleznyov, 2019; Willems & Van den Bossche, 2019). We also excluded conceptual and theoretical articles, and end-of-project reports. We found that some cross-cultural articles, that compare LS processes from different countries, did not describe each LS implementation with the same thoroughness. Since the analytic rubric that we designed for our assessment cannot account for this, we also excluded this type of study.

Thirdly, we concentrated our analysis on LS with in-service teachers. This criterion was at times difficult to assess, as some articles report on in-service teachers that are enrolled in graduate courses (e.g., Pang, 2016), or on graduate students conducting LS with a group of in-service teachers (e.g., Csida & Mewald, 2016). In order to systematize our decisions, we included articles that self-identify their teachers as in-service teachers, as well as articles that report on a LS group that consisted predominantly of in-service teachers.

Finally, we had intended not to exclude any studies based on quality criteria as long as they were peer-reviewed. During full-text screening, however, we struggled to fully understand five eligible articles. While these studies provided key words that seemed relevant to our analysis, they did so in inconsistent ways that challenged the coders' reliable and fair assessment. We therefore excluded these five studies on basis of their intelligibility. As specified in the pre-registered review protocol (Kager et al., 2021), we also excluded books and gray literature, as well as articles written in a language other than English, due to the authors' own language capabilities.

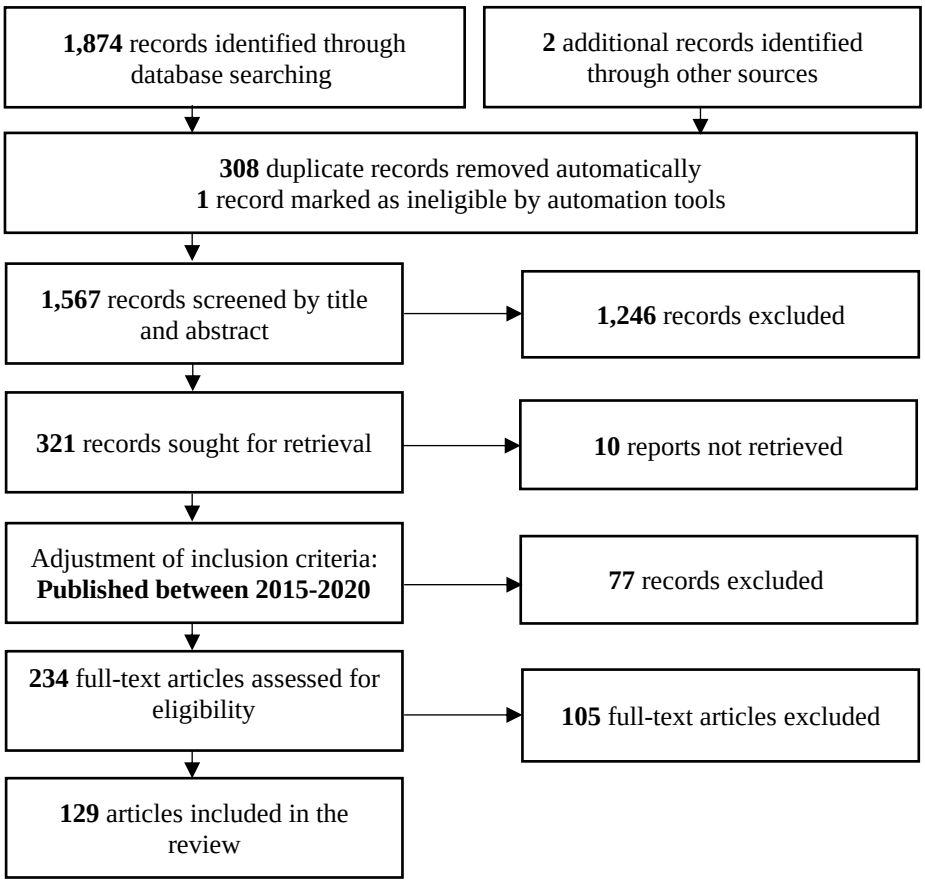
Search Procedure and Identification of Studies

The literature search comprised several stages. Firstly, we searched the databases SCOPUS, ERIC, PsychInfo, Academic Search Premier, Bibliography of Asian Studies, JTSOR, and ProQuest for articles published between 2010 and 2020 that included the term "lesson study" in their title, abstract, or key words. In later stages, we identified two additional records through referential backtracking. Altogether, the search yielded 1,876 records, which were imported to the reference management software Zotero. After the automatic and manual removal of duplicates (*N* = 308) and records that had been retracted after publication (*N* = 1), we

imported the remaining 1,567 records to Rayyan, a web-tool that supports the screening of literature (Ouzzani et al., 2016). The first author and a trained research assistant independently rated each abstract according to the set inclusion criteria. Disagreements on potential relevance of studies were discussed and solved collaboratively. The Rayyan' app gave the raters a near perfect intercoder reliability of above 95%. This step reduced the set to 321 articles.

Figure 1

Flowchart of the Literature Search Process Adapted from Moher et al. (2009)



Note. Records identified per data base: ERIC *N* = 913; Scopus *N* = 803; APA PsychInfo *N* = 100; Academic Search Premier *N* = 5; Bibliography of Asian Studies *N* = 5; JTSOR *N* = 7; ProQuest *N* = 41.

The full papers of these studies were imported to Zotero, with 10 potentially relevant studies excluded on the grounds that the texts were not publicly available. During the full-text screening, we made two changes to the inclusion criteria. Firstly, for the reasons outlined above, we adjusted the date range to only included articles published between 2015 and 2020. This led to the exclusion of 77 full texts. Secondly, we added inclusion criterion g, which specifies that the article had to focus on either the whole LS process or specifically on the observation and reflection stages. This criterion was added in order to ensure that all included studies could be expected to include relevant information about the observation and reflection stages. Overall, we excluded 105 studies in this phase. This left 129 studies in the review, which were subsequently coded in Excel. Figure 1 illustrates the stages of this culling process.

Data Coding and Analysis

The data coding and analysis followed five stages: 1) identifying categories; 2) developing the coding tool; 3) coding and assessing intercoder reliability; 4) extracting data of theoretical frameworks; and 5) data analysis.

Identifying categories. We began by reviewing the LS literature to identify a list of decisions taken by researchers and/or teachers that relate to the observation and reflection stages in LS. We piloted this list of categories by coding 25 randomly selected articles on LS. This took place before the systematic literature search and the piloted studies were not subject to our inclusion criteria. Based on our findings from the pilot coding, we refined the list and settled on eight categories for the assessment of transparency (Table 1). Each of these categories were assigned to one of Moravcsik's (2020) three dimensions of transparency.

Table 1

The Eight Categories Derived for the Assessment of Transparency

Level	Category	Description
Data Transparency	Means of Data Collection	How did teachers collect data, and what type of data did they collect?
	Focus of Observation	What did teachers observe?
	Scope of Observation	Whom did teachers observe?
Analytic Transparency	Interpretive Process	How did teachers reflect on the data?
	Procedure/Structure of the Post-Lesson Discussion	Did the reflection stage follow a specific procedure or structure?
Production Transparency	Role of Outside Expertise	Were outside experts present and what role(s) did they play?
	Documentation of Reflection Stage	Did someone document the reflection stage, and if yes, in what form?
	Setting of the Post-Lesson Discussion	Where and for how long did the reflection stage take place?

Developing the coding tool. The design of the coding protocol is based on Hallinger's (2014) analytic rubric, which uses three levels of distinction (i.e., an article does not include information, includes partial information, includes detailed information). We developed definitions and anchor examples for each category and level. The final coding protocol (Table S1) consisted of three parts: a *Quick Critical Appraisal Checklist* that reiterated the inclusion criteria; *Additional Information*, in which coders recorded general characteristics of the article, such as its research design and the label used to refer to the reflection stage (e.g., "post-lesson discussion"); and the *Analytic Rubric*, which included the eight categories outlined above for assessing transparency (see Table 1).

Coding and assessing interrater reliability. We coded the articles from the final set of studies according to a procedure adapted from O'Connor and Joffe (2020). The first author, who acted as the primary coder, coded a small amount of data during the development stage of the coding protocol to ensure its suitability. The first author then trained two research assistants by triple-coding studies, comparing results, discussing ambiguous examples and refining definitions in the coding tool. Satisfactory reliability was achieved after four rounds

of coding and an updated version of the coding protocol was established. Subsequently, the three coders worked independently but met regularly to discuss problematic cases. During this stage, the coders collaboratively assembled a list of keywords for each category (i.e., words associated with the reporting of a certain category, see Table S2). After the completion of the coding, we used the search function and the list of keywords to double check categories which we had rated with 0 (i.e., no information provided) to ensure that we had not missed any information. The first author double coded 20% of all studies, which has been suggested as an appropriate proportion for large data sets (O'Connor & Joffe, 2020). In order to account for the multiple coders, the first coder randomly selected and coded studies from each additional coder. Intercoder reliability (Table 2), calculated in R (R Core Team, 2013), was strong (McHugh, 2012).

Table 2

Overview of Intercoder Reliability across the Three Coders

	Studies coded (*N* = 129)	**Studies double coded by Coder 1** (overall 20%)	**Cohen's** κ
Coder 1	78	-	-
Coder 2	25	12	.80 (strong agreement)
Coder 3	26	13	.81 (strong agreement)

Extracting data of theoretical frameworks. After coding was completed, each coder searched their allocated articles for any theoretical frameworks on observation and reflection. This process was also supported by the list of keywords. Findings were recorded in the form of notes in Excel.

Data analysis. Finally, we recorded our findings in an overview Excel sheet to organize the information and calculate frequencies. The terms used in articles to refer to the reflection stage had to be organized in thematic groups in order to be quantified. We imported the list of all labels extracted from the studies to MaxQda (VERBI Software, 2019) and created a Code Co-occurrence Modell with MaxMaps. We first grouped the labels according to themes and developed codes, such as "discussion" and "conversation." To represent variations of the same concept (e.g., "reflection," "reflecting," and "reflective"), we grouped some words

under a joint label ("reflect*"). We double-coded labels that included several themes. For example, the label "post-lesson reflection" was double-coded as "post-lesson" and "reflect*," and the label "reflective debrief" as "reflect* and "debrief*."

For the analytic rubric, we calculated raw frequencies and percentages for each category as well as the total score for each article in Excel. These frequencies were imported to R (R Core Team, 2013) to create graphs. We then selected several qualitative examples and quotations from the reviewed studies to illustrate our findings and complement the analysis.

Results

The results are organized into three main sections. We first describe the general characteristics of the studies included in this review and then report the findings on the transparency of the LS intervention. Lastly, we present the findings on the theorization of the observation and reflection stages.

General Characteristics

The 129 studies included in this review took place in 33 countries. The countries represented with the highest numbers of studies are the US (*N* = 24), Turkey (*N* =13), and the United Kingdom (*N* = 12). As Table 3 indicates, the number of publications being published on LS has increased relatively consistently over the five years from 2015 to 2020.

Table 3

Number of Publications per Year and According to Region

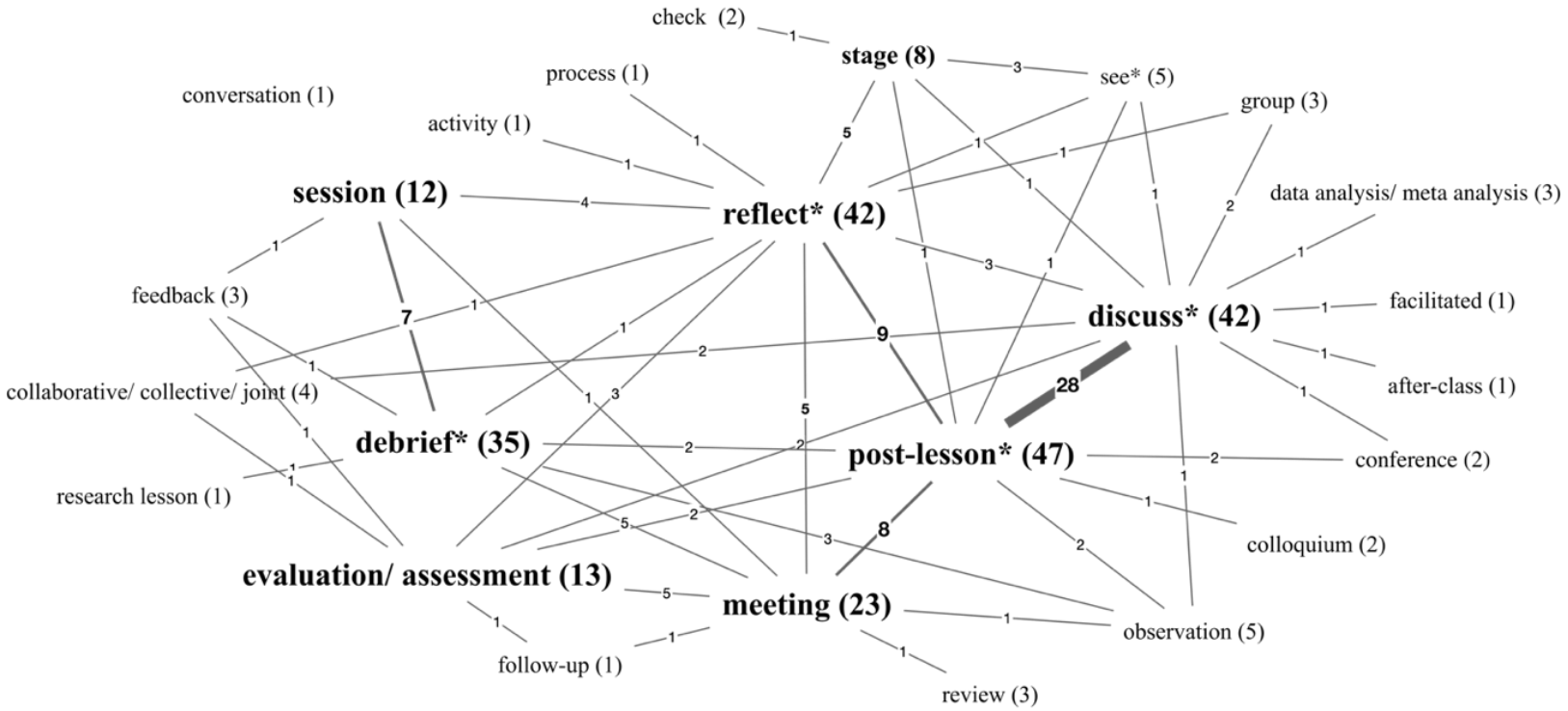
	2015	**2016**	**2017**	**2018**	**2019**	**2020**	Total
Asia	7	10	6	9	14	16	62
Europe	2	4	5	5	5	11	32
North America	5	5	4	5	4	5	28
Australia		1	1		1		3
Africa				1	1	1	3
South America						1	1
Total	14	20	16	20	25	34	129

The majority of studies described their PD approach as LS (*N* = 79), Japanese LS (*N* = 12), or Chinese LS (*N* = 6). Some studies used modifying words (i.e., participatory LS, blended LS), and three studies used an established acronym to refer to their LS adaptation, such as CLR (i.e., Collaborative Lesson Research). Most studies were conducted either in secondary school (*N* = 61) or primary school (*N* = 44). Almost half of all studies (*N* = 60) reported using some sort of case study design as their research methodology. Forty-nine studies reported that they employed a type of qualitative research design without further specifying their approach. Detailed tables for these general characteristics are included in the supplemental materials for this article (Tables S3–S6).

We documented a wide array of labels used to refer to the reflection stage. We also found variation within articles, with 25 studies using at least two different labels to refer to the reflection stage within the text. However, 15 studies did not make use of any specific label at all. The map in Figure 2 illustrates how often terms occurred by themselves or were used in combination with one another. The largest group consists of the phrase “post-lesson” (*N* = 47), followed by “discuss*” (*N* = 42) and “reflect*” (*N* = 42). The map also demonstrates that the by far most common combination was “post-lesson discussion” (*N* = 28), followed by “post-lesson reflection” (*N* = 9).

Figure 2

Co-occurrence Map: Labels, Words, and their Combinations Used to Refer to the Reflection Stage



Note. The number in brackets indicates the frequency of the word or phrase in the reviewed articles. The number on the connecting line indicates how often words or phrases were used in combination with each other within a phrase.

Assessment of the Transparency in the Observation and Reflection Stages

In this section we report the results of the analytic rubric, which was used to assess the transparency of articles when reporting the observation and reflection stages of LS. We will first present an overview of the total scores and then address each category individually.

Overall Rating

Figure 3 presents the distribution of the 129 studies included in this review by scores on the analytic rubric measuring eight categories for transparency. The categories were assessed with scores of 0 (does not include information), 1 (includes partial information), and 2 (includes detailed information). The maximum score would yield a rating of 16. The highest rated article scored 13 points (Aydogan Yenmez et al., 2017a), followed by two articles that scored 12 points (Færøyvik Karlsen, 2019; Warwick et al., 2016). Almost 50% of articles were scored between 6 and 9 points, the most frequently scored rating being 8 (*N* = 17). On the lower end, several articles met almost none of the eight criteria, with 21 articles scoring 2 or lower.

Figure 3

Distribution of Studies by Scores on the Analytic Rubric Measuring Eight Categories of Transparency

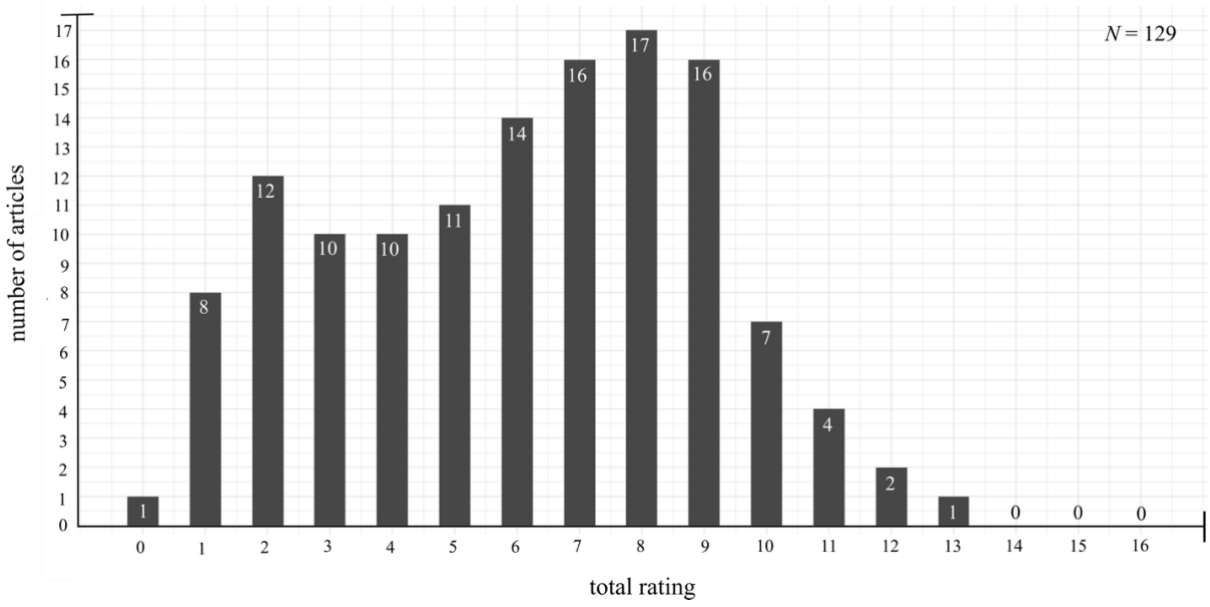
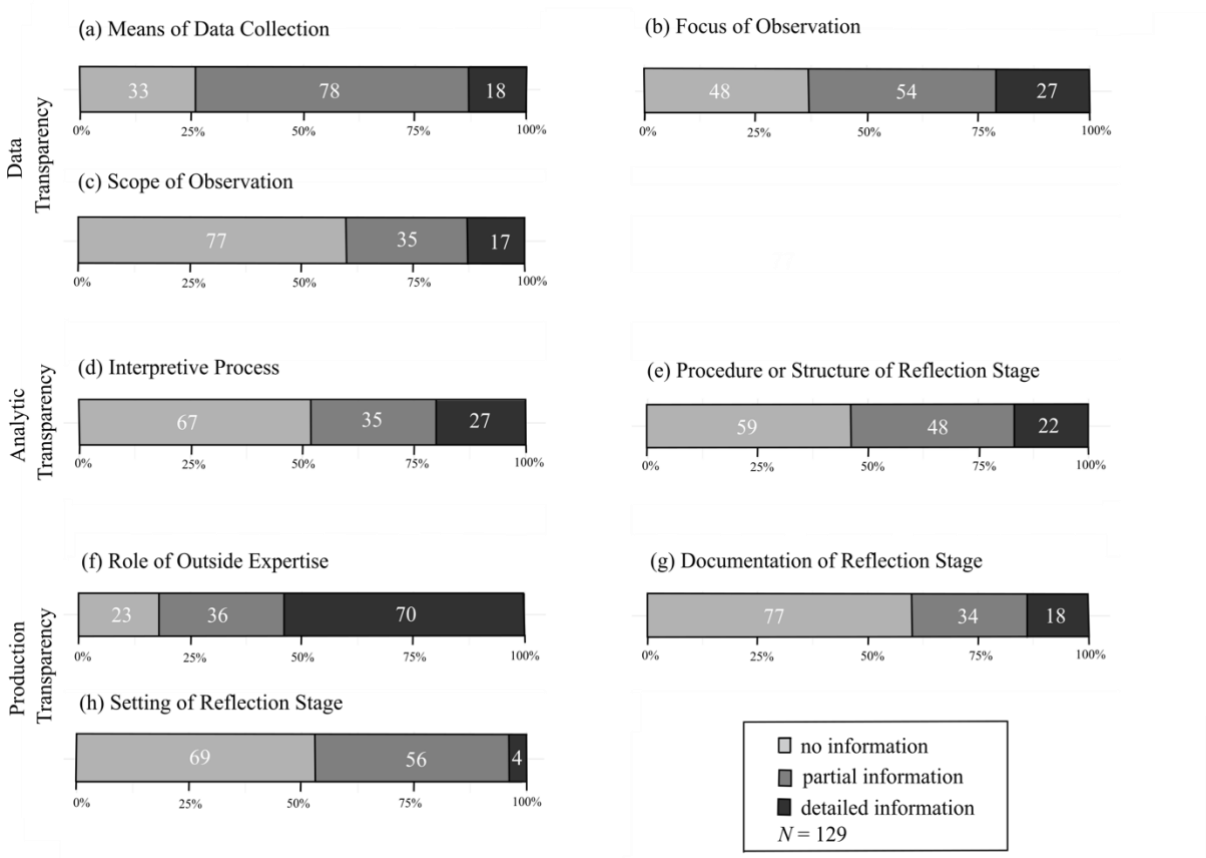


Figure 4 displays the assessment of transparency according to each category and indicates stark differences between the categories. The category *Role of Outside Expertise* was the most transparently communicated category by a large margin. Some categories, such as *Scope of*

Observation, Interpretative Process, Documentation of Reflection Stage, and *Setting of Reflection Stage*, were rated with 0 across the majority of the articles. In the following, we will discuss each category separately.

Figure 4

Assessment of Transparency According to the Eight Categories Defined in the Analytic Rubric



Note. Numbers inside bars represent raw counts.

Means of Data Collection

The majority of studies (61%) included some information on the type of data collected by teachers. The most common type of data was notes (*N* = 40), followed by videos or audio-visual recordings (*N* = 19), and student work (*N* = 17). A complete list of data types is presented in Table 4. Only a few articles (13%) explained the rationale behind the means of data collection or provided additional information about the process. Articles that did include this information described, for instance, that LS groups developed their own observation forms or rubrics (e.g., Bruce et al., 2016; Craney et al., 2020), or referenced existing templates or

material from a specific LS handbook (e.g., Khokhotva & Elexpuru Albizuri, 2019; Lucenario et al., 2016). A list of all articles rated with 2 for this category and their approaches to data collection can be found in the supplemental materials (Table S7).

Challenges encountered with this category. Several articles briefly mentioned "notes," "systematic observations," or "field notes from lesson observation," but failed to unambiguously state whether these notes had been taken by teachers, facilitators, or researchers, and whether these notes were analyzed by teachers during the reflection stage, rather than by researchers as part of their research study. Other articles mentioned LS handbooks or work-books, but did not specify which they were or provide any references.

Table 4

Type of Data Collected by Teachers during the Research Lesson

Type of data	*N*	Type of data	*N*
Notes	40	Private memos/reports	2
Videos, audio-visual Recordings, photos	20	Assessment instrument	1
Student work /artifacts	17	Rubric	2
Student interview	13	Notes on mobile phones	1
Observation Form/sheet/tool/template	8	Written feedback from students	1
Pre-post test	8	Blackboard writing	1
Observation protocol/ log/ Notebook	8	Articles stating that Observation followed no protocol	1
Lesson plan	4	No information given in article	33

Note. A total of 38 articles described collecting several types of data.

Focus of Observation

The majority of articles provided some information (43%) or even detailed information (21%) on the focus of teachers' data collection during the research lesson—that is to say, *what* teachers observed. For example, Won (2017) explains how teachers discussed the focus of their observations in the planning process and noted down expected or desired student responses in the lesson plan to guide their observations. Gilissen et al. (2020, p. 1261) describe teachers focusing their observations on students' ability for systems thinking in biology education. During the research lesson, teachers observed how the case students behaved,

communicated, and performed during certain key activities in the lesson, which are also detailed in the article.

Challenges encountered with this category. Studies that did not include any explicit information on the focus of teachers' observations usually mentioned student learning as the general focus of LS at some point in the article. We consider the notion of student learning, in this context, as vague and nondescriptive, as the term could potentially refer to almost any pedagogical activity that occurs in the classroom. The variety of observation foci that we found among the reviewed studies demonstrates the fact that only being given the information that teachers focused their observations on student learning is not sufficient to understand how this part of LS was executed and, as such, is certainly not replicable.

Scope of Observation

The majority of articles (60%) did not included information on whom teachers observed during the research lesson. Only 13% of articles provided explicit and detailed information on this topic. One example comes from Norwich et al. (2016, p. 183), who specify that each LS group in their study chose two students for observation. Teachers in this study based their selection on learning performance, observing both a student who usually struggled with the lesson's content and a student who represented a level that teachers felt was "typical" for this class. Liu (2016, p. 106), on the other hand, tells us that the teacher who implemented the lesson asked the other team members to form groups and each observe a subgroup of students. Their goal was to learn something about each student.

Challenges encountered with this category. The information about whom teachers observed was sometimes disclosed between the lines. For example, some articles mentioned at some point the number of students in the class, inviting the conclusion that teachers observed all students. The majority of articles did not, however, communicate this in an unambiguous way that did not require the reader to make any inferences. Furthermore, most articles that focused their analysis on student work rather than observational notes did not indicate whether or not the work of all students was considered in the reflection stage, or rather just the work of specific students.

Interpretative Process

About half of the articles (52%) did not clearly explain how teachers analyzed and reflected on the collected data. Twenty-seven percent of articles included partial information, that is, they provided examples of, or original quotations from, the interpretative process. And 21% of the articles included a definition or conceptualization of the reflection stage or of teachers' interpretative processes. Some of these articles did so in passing, while others dedicated more time to the issue.

Challenges encountered with this category. A variety of studies briefly referred to concepts or terminology in connection to the reflection stage of LS. These articles did not, however, provide a definition or explanation for the relevant terms. Similarly, some articles mentioned reflection in connection to concepts such as the community of inquiry or professional learning communities, but did not explicitly conceptualize or define reflection itself.

Procedure/Structure of the Reflection Stage

Almost half of all studies (46%) did not specify how the reflection stage of LS was structured, specifically whether or not teachers followed a specific procedure. Only 17% of articles provided detailed descriptions of these processes. These usually included a chronological component. For example, Huang et al. (2017) relate how teachers first shared their reactions to the lesson, discussed the learning outcomes, and then talked about their concerns. Kanellopoulou, and Darra (2018a, p. 71), on the other hand, describe teachers following a research lesson review protocol adopted from Stepanek et al. (2007), and list several chronological steps followed by these teachers.

Challenges encountered with this category. Several studies that were coded as providing no information on this step did still include some indication of what a typical post-lesson discussion might include in a general sense. We usually found this information in the studies' literature review when the specifics of LS were introduced. These studies did not, however, define what their own implementation of LS looked like and they failed to clearly state whether or not their adaptation included any or all of these typical steps, and in what order those steps were taken.

Role of Outside Expertise

With only 18% of articles not including this information, this category was communicated in a largely transparent manner in most studies and the majority of articles included detailed information on the roles outside experts took. Pang (2016), for instance, reports on LS based on the collaboration between a university professor and in-service teachers at a Korean primary school who were enrolled in a graduate course. Pang informs the readers that she took on the role of the "knowledgeable other" and shared her expertise with the LS group, for example by commenting on the lesson plans and providing feedback during the reflection stage.

Challenges encountered with this category. The analysis indicated that researchers take on a variety of roles in the LS process. Pang (2016), for example, explicitly states that researchers acted as external facilitators, providing expertise and guidance to the LS group. The studies by Norwich et al. (2016, 2018) report that the LS group was joined by both the researchers and additional experts. Some researchers accompanied the process as active participants in the LS process and simultaneously acted as authors of the research paper (e.g., Leong et al., 2016; Ni Shuilleabhain & Seery, 2018), while others described their role as researchers being that of invisible observers (Moghaddam et al., 2015). This diversity made it difficult to clearly understand the role of researchers and external experts in articles that mentioned external instructors or experts, but neither identified them nor explained their role in the LS process.

Documenting the Reflection Stage

The majority of articles (60%) did not include any explicit information about whether someone documented the group's reflection process and their take-aways in any way. Only 14% of articles provided detailed information about this. Watanabe et al. (2019), for example, include an appendix with documentation from the LS process that could serve as templates for others. Celik and Güzel (2018, p. 182) describe teachers keeping individual reflective diaries after each LS cycle to record their experiences and thoughts in regard to specific questions they faced. Another example comes from Moss et al. (2015), who report teachers documenting the LS process in a so-called "LS package and iBook," which can be accessed online by anyone interested in learning more about their study.

Challenges encountered with this category. Several articles referred to notes or records but failed to clearly describe who took those notes and at what point in time, nor even whether the purpose of the notes was to document the LS process.

Setting of the Reflection Stage

This category examined whether articles included information on the duration of and/or setting for the reflection stage. The majority of articles (54%) did not include any explicit information on this. Across the remaining articles, 28 included details on the length of the reflection stages, the most common duration being one hour (*N* = 8), followed by up to one hour (*N* = 7), up to two hours (*N* = 7), and longer than two hours (*N* = 6). Concerning timings, articles usually specified whether the reflection stage had taken place immediately after the research lesson, or some time later. Bradshaw and Hazell (2017), for example, report that the teachers' reflection stage followed soon after the teaching session so that "ideas and observations from the lessons were strong in the minds of the observers" (p. 34). Whereas Aydogan Yenmez et al. (2017a, p. 321) tell us that the students' reports—which were the basis for data analysis—were copied after the research lesson so that each teacher would have their own copy available to them during the reflection stage.

Challenges encountered with this category. This category was easy to code, as the vast majority of articles did not provide any information on this issue.

Theoretical Frameworks for the Observation Stage and Reflection Stage

Observation stage. We found 10 studies (8%) that explicitly connected the observation stage to a theoretical framework or to concepts of observation that already exist in the literature. Five of these articles referred to the notions of "(professional) noticing" and "professional vision" (based on e.g., Jacobs et al., 2010; Sherin & Han, 2004; Van Es, 2011). In Karlsen and Helgevold (2019), professional noticing was in fact the focus of their research objectives, exploring the depth of teachers' observations and their analytic stance in the post-lesson reflection. They conclude that teachers' professional noticing in LS should be supported by observation forms designed explicitly to capture student learning. Other articles referenced more general frameworks, such as active learning (Garet et al., 2001) or theories of teacher learning (Marton, 2015; Penuel et al., 2007), while explicitly highlighting observation and its role within these frameworks. Koutsouris et al. (2017) used Dyke et al.'s (2006) notion of

"tunnel vision" to elaborate on difficulties with videotaping the research lesson and to describe the effect classroom videos might have on its observers and those being observed. A list of these studies and their approaches can be found in the supplemental materials (Table S8).

Reflection stage. We found 20 studies (16%) that explicitly theorized teachers' reflection processes. In general, reflection was identified as an important aspect in teacher learning and several articles ground their understanding of reflection in the works of Dewey (1933) and Zimmermann (2000). The most frequently cited scholar was Schön (1983, 1995), with six articles referring to his notion of the reflective practitioner, as well as reflection-in-action and reflection-on-action. Another reoccurring framework was rooted in the theory of cognitive conflict (e.g., Limon, 2001; Piaget, 1985; Posner et al., 1982). A list of the studies that theorized reflection can be found in the supplemental materials (Table S9).

Alternative conceptualizations of the reflection stage. In addition to the frameworks discussed above, we found that seven studies (5%) grounded their understanding of the reflection stage in alternative theoretical perspectives. Brown et al. (2016), for instance, referred to theoretical perspectives on "learning conversations," and Lee and Tan (2020) on "professional conversations." Warwick et al. (2016) and Bae et al. (2016) both connect the reflection stage to the notions of dialogue, interthinking, and modes of talk (Littleton & Mercer, 2013; Mercer, 2000). A list of these studies and their approaches can be found in the supplemental materials (Table S10).

Discussion

This systematic review set out to examine two research questions. Firstly, we asked how transparent in-service LS articles are in reporting on their observation and reflection stages of LS. And, secondly, we asked which theoretical frameworks are currently being used to conceptualize these two stages. In regard to the first question, our analysis of 129 articles indicates that several categories across all three dimensions of transparency (Moravcsik, 2020) were either omitted completely or described only partially in the majority of studies. In line with Cheung and Wong (2014) and Larssen et al.'s (2018) previous assessments, these findings provide broad evidence of a lack of transparency on two crucial stages of the LS interventions in the current LS literature. In regard to the second question, we discovered that only a small minority of studies theorized the observation and reflection stages of LS. These

findings are also consistent with previous assessments from the field of education and social sciences (DeLuca et al., 2015; Mann & Walsh, 2013), reporting that frequently used concepts, such as reflection, often remain undertheorized in publications.

The primary reason that transparency represents an issue for the current state of LS literature is that research studies that are not transparent about their actual methods and execution of their intervention may not be particularly helpful for researchers and practitioners. When studies omit procedural details about the intervention then researchers and teachers are not able to fully comprehend the interventions nor utilize their outcomes. We now want to look at three aspects of this issue that emerged from our analysis to discuss in more detail: the omission of information; LS as both a research method and a research object; and the lack of a shared theoretical framework for LS.

Omission of Information

The analysis found that, in the studies we looked at, some categories of transparency were communicated more clearly than others due to information being omitted. In general, the reasons for this omission appeared to be the presumption of a shared understanding of LS, a lack of awareness that certain kinds of information might be important to understanding findings on LS, and an unbalanced focus on LS outcomes over LS processes. We will discuss these issues using the following examples.

The category *Role of Outside Expertise* was the most transparently communicated category and is also frequently the subject of research (Amador & Carter, 2018; Hauge, 2021; Lewis, 2016; Takahashi, 2014). This scientific discourse seems to be driven by researchers' own interest in how to best initiate, lead, and sustain LS. As our review showed, the researchers writing the studies are frequently also personally involved in LS as project leaders, coordinators, and educators. We assume that the researchers' active roles in LS have translated to the high transparency in the communication of this role in our findings. Consequently, a shared understanding of the multifaceted roles of outside expertise is openly discussed and the importance of this information seems to be recognized in the majority of the publications reviewed.

This shared understanding concerning a certain part of the LS intervention—and its importance to LS—was largely lacking for the remaining categories assessed in this review. Four categories (*Interpretative Process, Scope of Observation, Documentation of Reflection*

Stage, and *Setting of Reflection Stage*) were scored "not included" in the vast majority of articles. If mentioned at all, these categories were frequently described in vague or general terms that left crucial details out, such as how teachers carried out the reflection process. The low transparency of these categories might stem from researchers' belief that it is enough to indicate how LS is "commonly" or "usually" conducted. This belief, however, is refuted by the abundance of distinct approaches to this stage described in a number of the articles assessed in this review. For example, the 18 articles that clearly communicated how LS groups documented their learning all differed in how this was conducted. The approaches included reflective diaries or journals kept by teachers (Calleja & Formosa, 2020; Çelik & Güzel, 2018), a specific template for note taking (Lee & Tan, 2020), meeting calendars (Kanellopoulou & Darra, 2018a), a report prepared by either the whole LS group (Özdemir, 2019) or a designated group member (Chua, 2019), as well as teachers' individual documentation of the process in an online space (Joubert et al., 2020). The way a LS process is documented can play an important role in structuring the reflection stage (Kager et al., 2022), but it also facilitates the ways in which groups mobilize and share their knowledge. This is considered a crucial part of Japanese LS and important to sustaining LS in schools, but is frequently neglected in LS translations (Seleznyov, 2018). This abundance of ways in which LS groups document and mobilize their learning demonstrates that the research community cannot and should not presume that there is a standard process of documenting teachers' learnings in LS that requires no further communication in research articles. Only by explicitly reporting details about these steps can the process of creating a common understanding about these aspects be advanced.

Another source of low transparency was the predominance of articles reporting LS outcomes over its processes. This underreporting of information concerning the production of research, or in this case an educational intervention, can stem, for instance, from researchers preferring a clear "storyline" over descriptions of trial and error (Aguinis et al., 2018), or—especially in qualitative research—from trying to keep to strict word limits imposed by journals (Moravcsik, 2020). These abridged descriptions sideline valuable information about judgement calls and choices crucial if others are going to be able to replicate the research study or intervention in question (Aguinis et al., 2018). Focusing publications on reporting the findings over how they were generated can make it difficult for other researchers to understand and evaluate the meaning and value of the research. In addition to

this, steps that appear trivial to some may be valuable to others wanting to improve their own LS practice. Aydogan Yenmez et al. (2017a, p. 321), for example, specify that teachers prepared and handed out copies of student work before the start of the reflection stage. This small detail can serve as a practical tip to educators using or considering introducing LS in their own schools. Our analysis suggests that even just a short statement or description of such details can greatly enhance the communication of the concrete details of how LS was implemented, which would be of benefit both scientifically and pedagogically.

Several of the articles reviewed did, however, provide innovative solutions for the problem of strict word limits and restrictive formats. By including links or references to supplementary materials stored on journal websites, online repositories, or school- or project-specific websites, these articles found an effective way of making their materials widely accessible to others. Sharing data and materials openly in order to enhance transparency is central to the Open Science movement (Nosek et al., 2012), which is becoming increasingly important in educational science (van Dijk et al., 2021). Our review suggests that Open Science practices can also advance and deepen discourses in the field of LS.

Is Lesson Study a Scientific Method, Teacher-Led Research, or a Research Object?

Our findings show that the transparency of the articles we reviewed was further complicated by the fact that LS was approached in quite different ways: as a scientific research method, as teacher-led research, and as a research object. Some articles stated that LS itself was used as a research method by researchers (akin to action-based research) to explore, for example, how to best teach fractions. Researchers therefore conducted research *through* LS, rather than *on* LS. Other articles conceptualized LS as teacher-led research, with the researcher(s) taking on an active part in the LS group and frequently focusing their articles on relating their experiences. The vast majority of articles, however, viewed LS as a teacher-led PD approach (i.e., an intervention) and research object that was investigated through the use of a separate methodology, such as a case study approach or design-based research. This last type of LS research can arguably produce the most trustworthy, replicable, and comparable evidence for the efficacy of LS or its' use in distinct contexts, as a scientific research method is employed to conduct research *on* LS.

Regardless of whether LS is viewed as a research method, teacher-led research, or a research objective, articles should always adhere to the principle of research transparency,

that is, to clearly report their evidence, analysis, and overall research design (Moravcsik, 2020). Part of this is to describe the LS intervention in enough detail so that others can understand or replicate it. Our findings indicate that the majority of articles did not provide such descriptions regarding the observation and reflections stages of LS and did not adhere to any discernable reporting standard. In fact, information related to the LS intervention was sometimes reported in unexpected places, such as the theory section or discussion. Other articles scattered the information across multiple sections, with relevant information sometimes appearing only late in the text. Another problem was that articles that conducted research on LS frequently failed to separate the descriptions of their research method from those of the LS intervention. Some articles, for example, reported the data collected by both researchers and teachers in the same chapter, sentence, or even bullet list, making it unclear who had collected which data for what purpose.

In order to avoid confusion, we recommend that articles clearly position themselves as either research *through* or *on* LS. Research *on* LS needs to clearly separate descriptions of their research method from descriptions of the LS intervention. We suggest to report the LS intervention in a separate subchapter within the method section. It is further important to use unambiguous terminology. For example, if both researchers and teachers collected observational notes during and of the LS process, these different types of notes need to be clearly identifiable through the use of consistent language.

Lack of a Shared Theoretical Framework

Another source of low transparency in our sample was the frequent use of the terms ‘observation’ and ‘reflection’ without providing clear definitions or situating these constructs within a theoretical framework. In fact, only a small minority of articles clearly defined relevant terms, and 92% and 79% of articles undertheorized the observation and reflection stages, respectively. These findings provide new and concrete insights into previous assessments of the level of undertheorization in LS research (Elliott, 2012; Stigler & Hiebert, 2016) and demonstrate that the LS community uses a diverse set of terminology and labels while assuming that there is a shared understanding of these concepts. As our findings demonstrate that this shared understanding cannot be guaranteed, the lack of definitions and theorization renders terms such as ‘observation’ and ‘reflection’ untransparent in LS research.

Several researchers have shown that the undertheorization of concepts is a reoccurring problem across the social sciences (Fleetwood & Hesketh, 2006; Radovic et al., 2018; Wang et al., 2020; Wolgemuth et al., 2017). Empirical research that concentrates on practical descriptions, such as LS research, seems particularly susceptible to this problem (Fleetwood & Hesketh, 2006). As Fleetwood and Hesketh (2006) argue, missing theorization raises concerns around whether or not results can be sufficiently accounted for, and whether practices are recommended based on confirmed relationships. Lewis et al. (2006) also point toward this concern by referring to a "lack of clear causal warrant" (p.7) in the field of LS. In other words, the high contextualization of LS makes it difficult for researchers to identify which practices cause certain outcomes. Routes to circumvent this shortcoming of LS research include explicit descriptions of local LS interventions (Lewis et al., 2006), and, as we argue, a stronger theoretical footing to provide explanatory power and guidance as to how teachers' observation and reflection processes can be structured and explained.

Based on our findings, we can see at least two explanations for and consequent challenges of the undertheorization of concepts in the practice-oriented field of LS. To begin with, there is an absence of standardized procedures for the observation and reflection stages (Cerbin & Kopp, 2006), and, as our analysis suggests, of a consistent terminology to talk about these stages. This is illustrated, for example, by the number of different labels used to refer to the reflection stage of LS (Figure 2). With a lack of a shared foundation, these labels, which are predominantly not theoretically informed in the articles we reviewed, pose several critical questions: Do teachers pursue the same goal in a "debrief," a "post-lesson reflection," and a "data analysis"? Does the inconsistent terminology suggest different ways of implementing the reflection stage? And in what ways does the implementation matter to the LS outcome and teachers' subsequent instructional improvements? In order to answer the last question, which arguably represents one of the most essential critical research objectives in the field (Lewis et al., 2006), it seems clear that we need to strive for greater conceptual coherence in LS studies in order to establish a common point of departure.

Secondly, LS is not rooted in a specific theory, but makes use of theories generated or developed in other fields. Lewis et al. (2019) explored how self-determination theory, self-efficacy theory, and knowledge integration theory can all inform research on LS. Empirical studies frequently underpin their LS research with models of PD and teacher growth, such as those by Clarke and Hollingsworth (2002) and Guskey (2002), or, in the case of Huang et al.

(2016), they develop their own theory-based LS model. We have also seen the generation of new theories from empirical research on LS, such as Mynott's (2019) theoretical outcome model of LS. In addition to this, this review identified a number of useful theories for the conceptualization of teachers' learning processes in LS, such as cognitive conflict, modes of teacher talk, and professional noticing. Some of these approaches have been picked up and further investigated in recent studies, such as Dick et al. (2022), Hrastinski (2021), and Karlsen and Ohna (2021) for the professional noticing of teachers, and Uştuk and De Costa (2021) and Kager et al. (2022) for critical and collaborative reflection. This development indicates that these theoretical perspectives are being actively explored and tested for how suitable they are for explaining the processes behind LS.

The transference of theories to new fields is a common practice, but brings with it a range of challenges, as theories might fit to some, but not all, aspects of the new context (Wang et al., 2020). In the case of LS, it seems that theorization does exist and approaches, perspectives, and models have been applied, yet they have neither been sufficiently advanced within the field nor adopted by the broader research community in their empirical research. As a symptom of both the undertheorization and the adoption of LS to new cultural contexts, a complex web of ambiguous terminology has developed. We argue that the field of LS has advanced to a point at which it would benefit from some standardization in order to negotiate what Kim (2021) refers to as a "conceptual grid" for LS outside of Japan. Importantly here, we are not suggesting standardizing LS as an intervention, but rather standardizing the way we talk about it.

A Framework for Reporting the Observation and Reflection Stages

We want to conclude our review by making the following recommendations concerning the reporting of LS interventions in research publications (Table 5). Firstly, researchers should aim to communicate their specific LS intervention in a concise way within the article, such as a subchapter as part of the method section. Secondly, researchers should strive to employ clearer terminology. This means that the specific use of terms such as "observation" or "reflection" need to be explained, and ideally derived from or embedded in a theoretical framework. It also means that researchers should be aware that, without sufficient explanation, readers are likely to draw their own conclusions concerning terminology or labels used in the text. Thirdly, we recommend the use of the following checklist based on the

findings of this review. The checklist can be used by researchers to evaluate the transparency of their manuscript and decide which aspects of their LS intervention need to be communicated to guarantee the usability of their research.

Table 5

Checklist of Items Recommended for Inclusion when Reporting on the Observation and Reflection Stages of Lesson Study

LS Stage	Checklist Item
Observation Stage	
Theoretical framework	How did researchers (and the LS group) understand the observation process from a theoretical perspective?
Type of data	What kind of data did teachers collect (e.g., structured notes, videos, student work…)?
Process of data collection	How did teachers collect this data? Was data collection guided by a specific protocol?
Focus of observation	What did teachers focus on in their observations (e.g., which aspect of student learning)?
Scope of observation	Did teachers observe the entire class, a subset of students, or specific students? What guided this decision?
Outside expertise	How were outside experts involved in the observation stage?
Materials	Can materials used in the observation stage be accessed elsewhere?
Reflection Stage	
Theoretical framework	How did researchers (and the LS group) understand the reflection process from a theoretical perspective?
Process of reflection	How did teachers carry out the collaborative reflection?
Structure of reflection	How was the reflection stage structured chronologically and what activities were involved?
Length of reflection stage	How long did teachers reflect together?
Setting of reflection stage	How was the reflection stage influenced by other contextual factors or decisions (e.g., time, space, …)?
Outside expertise	How were outside experts involved in the reflection stage?
Documentation	How was the reflection stage (or LS process) documented?
Materials	Can materials, such as reflection protocols, be accessed elsewhere?

Limitations

Our methodology is subject to a number of limitations. Firstly, it is important to note that the study focused solely on how transparent descriptions of LS's observation and reflection stages were. We recognize that a multitude of additional factors contribute to an LS outcome, including social and cultural contexts, hierarchical structures within the LS groups, the groups' motivation, and teachers' experience (Bocala, 2015; Hadfield & Jopling, 2016; Seleznyov et al., 2021). Secondly, we did not assess an article's quality or overall research transparency, but the degree of transparency with which an article communicated the observation and reflection stages of LS. The total rating given to an article does therefore not provide any assessment about the overall quality or scientific value of the article. Thirdly, we treated all categories assessed in the analytic rubric equally in our analysis and did not assign any weight to them. This choice might skew the results in so far as not all categories are likely to have the same impact on the outcome of an LS cycle. In order to assign weight to the categories, however, we would need further research that can provide a justification for this weighting. We would like to propose this as an avenue for future research.

There might be additional eligible articles that were not included in this review, as no database has complete coverage. Likewise, the list of categories assessed in this review were derived from the research literature, yet there might be additional categories of interest that we did not cover. The assessment of transparency, while guided by an analytic rubric, demanded definitive choices by the coders. These choices were not always easy, as they required coders not to try and read between the lines or make inferences. Nevertheless, we achieved high intercoder reliability and our findings are consistent with previous evidence. We have detailed further challenges that we faced in the assessment process in the findings section to enhance the transparency of this analysis process.

Lastly, we recognize Ishii's (2017) concern that research in LS frequently focuses on reflection at the expense of LS's first two phases—identifying a research question and planning instruction. Our analysis adds to this bias in so far as we only examine the observation and reflection stages of LS. We hope, however, that the present review can act as a springboard for future research into the transparency of each of the core stages of LS.

Concluding Remarks and Implications

The present review has confirmed that the field of LS is currently marred by low transparency in how the observation and reflection stages are communicated in the research. These findings build on similar observations about the undertheorization of LS by Cheung and Wong (2014), Stigler and Hiebert (2016), and Larssen et al. (2018). We broaden their analyses by outlining reasons for this undertheorization and subsequently recommending specific communication practices for empirical research on LS. The proposed checklist can, in the first instance, support practitioners in their implementation of LS and, in the second, motivate researchers to rigorously and comprehensively question and document their decisions on the implementation of LS, even when it appears trivial.

We draw a range of practical and theoretical implications from these findings. Our review underlines Lewis et al.'s (2006) argument that in order to make LS effective we need to identify its crucial underlying processes and implementation steps. We saw that explicit descriptions of the intervention can greatly contribute to the building of just such a knowledge base. In this sense, we hope that the lists and examples provided in the current review and its supplemental materials of articles that explicitly communicated their interventions can act as a resource on how to conduct and establish standards and on how to report the observation and reflection stages in LS. Our research further implies that Open Science practices, such as providing open access to resources and making data publicly available, can positively impact knowledge generation in the field of LS and ensure the usefulness and replicability of research.

Turning to theoretical implications, our review highlights the need for further theoretical development for LS in general, and the observation and reflection stages in particular. The theorization of these stages was almost always absent in the articles reviewed, though some articles presented promising avenues to stronger theorization. While a complete theory of LS might be too ambitious due to its variable and contextualized character, it does seem possible to advance these existing theories in the field of LS and to increasingly integrate them into empirical research in a more comprehensive, extensive, and thus potentially valuable way.

References

References marked with an asterisk indicate studies included in the systematic review.

*Abdella, A. S., Reddy, C., & Carl, A. (2018). Lesson Study in Eritrea: Its impact on middle school science teachers' learning and classroom practice. *Journal of Turkish Science Education*, *15*(3), 1–26. https://doi.org/0.12973/tused.10234a

*Abu-Alghayth, K., Jones, P., Pace-Phillips, D., & Meyers, R. (2020). Through the looking glass: Lesson Study in a center school. *International Journal of Educational Methodology*, *6*(2), 423–433. https://doi.org/10.12973/ijem.6.2.423

Aguinis, H., Ramani, R. S., & Alabduljader, N. (2018). What you see is what you Gget? Enhancing methodological transparency in management research. *Academy of Management Annals*, *12*(1), 83–110. https://doi.org/10.5465/annals.2016.0011

*Allan, D., O'Doherty, E., Boorman, D., & Smalley, P. (2020). Lesson Study and the construction of capital: Empowering children through dialogic engagement. *Education 3-12*, *49*(6), 730–744. https://doi.org/10.1080/03004279.2020.1781224

*Alwadi, H. M., Mohamed, N., & Wilson, A. (2020). From experienced to professional practitioners: A participatory lesson study approach to strengthen and sustain English language teaching and leadership. *International Journal for Lesson & Learning Studies*, *9*(4), 333–349. https://doi.org/10.1108/IJLLS-10-2019-0072

Amador, J. M., & Carter, I. S. (2018). Audible conversational affordances and constraints of verbalizing professional noticing during prospective teacher Lesson Study. *Journal of Mathematics Teacher Education*, *21*(1), 5–34. https://doi.org/10.1007/s10857-016-9347-x

American Psychological Association. (2022). *Publication manual of the American psychological association*. *1*, 1–xxii. https://apastyle.apa.org/products/publication-manual-7th-edition

*Anif, S., Sutopo, A., & Prayitno, H. J. (2020). Lesson Study validation: Model for social and natural sciences teacher development in the implementation of national curriculum in Muhammadiyah Schools, Indonesia. *Universal Journal of Educational Research*, *8*(1), 253–259. https://doi.org/10.13189/ujer.2020.080132

*Asari, S., Fauziyah, N., & Uchtiawati, S. (2018). Improving Teacher Pedagogic Competences in Remote Areas through Lesson Study Activity. *International Journal of Education and Literacy Studies*, *6*(2), 53–62. https://doi.org/10.7575/aiac.ijels.v.6n.2p.53

*Austin, L. (2017). Leading the introduction and development of Lesson and Learning Study in an English Secondary Academy. *International Journal for Lesson and Learning Studies*, *6*(1),

80–96. https://doi.org/10.1108/IJLLS-10-2016-0036

*Aydogan Yenmez, A., Erbas, A. K., Alacaci, C., Cakiroglu, E., & Cetinkaya, B. (2017a). Evolution of mathematics teachers' pedagogical knowledge when they are teaching through modeling. *International Journal of Education in Mathematics, Science and Technology*, *5*(4), 317–332. https://doi.org/10.18404/ijemst.296552

*Aydogan Yenmez, A., Erbas, A. K., Cakiroglu, E., Alacaci, C., & Cetinkaya, B. (2017b). Developing teachers' models for assessing students' competence in mathematical modelling through lesson study. *International Journal of Mathematical Education in Science and Technology*, *48*(6), 895–912. https://doi.org/10.1080/0020739X.2017.1298854

*Bae, L. C., Hayes, K. N., Seitz, J., O'Connor, D., & DiStefano, R. (2016). A coding tool for examining the substance of teacher professional learning and change with example cases from middle school science lesson study. *Teaching and Teacher Education*, *60*, 164–178. https://doi.org/10.1016/j.tate.2016.08.016

*Barber, K. (2018). Developing teachers' mathematical-task knowledge and practice through lesson study. *International Journal for Lesson and Learning Studies*, *7*(2), 136–149. https://doi.org/10.1108/IJLLS-09-2017-0042

*Baz, M. A. (2020). Improving visual literacy skills through news photos. *Journal of Inquiry Based Activities*, *10*(1), 61–82. https://ated.info.tr/ojs-3.2.1-3/index.php/ated/article/view/12

Bjuland, R., & Mosvold, R. (2015). Lesson study in teacher education: Learning from a challenging case. *Teaching and Teacher Education*, *52*, 83–90. https://doi.org/10.1016/j.tate.2015.09.005

*Bocala, C. (2015). From Experience to Expertise: The Development of Teachers' Learning in Lesson Study. *Journal of Teacher Education*, *66*(4), 349–362. https://doi.org/10.1177/0022487115592032

*Boran, E., Tarım, K., & M. Sencer Bulut Özsezer. (2018). The effect of Lesson Study on the perceptions of mathematics teachers about suject area competencies. *European Journal of Education Studies*, *5*(8), 371–394. https://doi.org/10.5281/ZENODO.2529510

*Bozkurt, E., & Yetkin-Özdemir, İ. E. (2018). Middle school mathematics teachers' reflection activities in the context of Lesson Study. *International Journal of Instruction*, *11*(1), 379–394. https://doi.org/10.12973/iji.2018.11126a

*Bradshaw, Z., & Hazell, A. (2017). Developing problem-solving skills in mathematics: A lesson

study. *International Journal for Lesson and Learning Studies*, *6*(1), 32–44. https://doi.org/10.1108/IJLLS-09-2016-0032

Brosnan, A. (2014). Introducing lesson study in promoting a new mathematics curriculum in Irish post-primary schools. *International Journal for Lesson and Learning Studies*, *3*(3), 236–251. https://doi.org/10.1108/IJLLS-09-2013-0050

Brown, C., Poortman, C., Gray, H., Groß Ophoff, J., & Warf, M. (2021). Facilitating collaborative reflective inquiry amongst teachers: What do we currently know? *International Journal of Educational Research*, *105*(101695), 1–11. https://doi.org/10.1016/j.ijer.2020.101695

*Brown, C., Taylor, C., & Ponambalum, L. (2016). Using Design-Based Research to Improve the Lesson Study Approach to Professional Development in Camden (London). *London Review of Education*, *14*(2), 4–24. https://doi.org/10.18546/LRE.14.2.02

*Bruce, C. D., Flynn, T. C., & Bennett, S. (2016). A focus on exploratory tasks in Lesson Study: The Canadian "Math for Young Children" Project. *ZDM: The International Journal on Mathematics Education*, *48*(4), 541–554. https://doi.org/10.1007/s11858-015-0747-7

*Bruce, C. D., & Hawes, Z. (2015). The role of 2D and 3D mental rotation in mathematics for young children: What is it? Why does it matter? And what can we do about it? *ZDM Mathematics Education (*, *47*(3), 331–343. https://doi.org/10.1007/s11858-014-0637-4

Bryk, A. S. (2015). 2014 AERA Distinguished Lecture: Accelerating how we learn to improve. *Educational Researcher*, *44*(9), 467–477. https://doi.org/10.3102/0013189X15621543

*Bussi, M. G. B., Bertolini, C., Ramploud, A., & Sun, X. (2017). Cultural transposition of Chinese lesson study to Italy: An exploratory study on fractions in a fourth-grade classroom. *International Journal for Lesson and Learning Studies*, *6*(4), 380–395. https://doi.org/10.1108/IJLLS-12-2016-0057

*Bütün, M. (2019). Mathematics teachers' early Lesson Study experiences in Turkey: Challenges and advantages. *World Journal of Education*, *9*(5), 51–62. https://doi.org/10.5430/wje.v9n5p51

*Cajkler, W., Wood, P., Norton, J., Pedder, D., & Xu, H. (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), 192–213. https://doi.org/10.1080/02671522.2014.887139

*Callahan, C. (2018). Promoting second-order historical domain knowledge through recursive and collaborative professional development. *The Journal of Social Studies Research*, *42*(3),

287–297. https://doi.org/10.1016/j.jssr.2017.06.001

*Callahan, C. (2019). Middle school geography teachers' professional development centered around historical photographs. *Journal of Social Studies Research*, *43*(4), 375–388. https://doi.org/10.1016/j.jssr.2018.11.003

Calleja, J., & Formosa, L. (2020). Teacher change through cognitive conflicts: The case of an art lesson study. *International Journal for Lesson & Learning Studies*, *9*(4), 383–395. https://doi.org/10.1108/IJLLS-05-2020-0028

*Cammarata, L., & Haley, C. (2018). Integrated content, language, and literacy instruction in a Canadian French immersion context: A professional development journey. *International Journal of Bilingual Education and Bilingualism*, *21*(3), 332–348. https://doi.org/10.1080/13670050.2017.1386617

*Canonigo, A. M. (2016). Using a non-coercive process to engage mathematics teachers in lesson study. *International Journal for Lesson and Learning Studies*, *5*(4), 329–347. https://doi.org/10.1108/IJLLS-02-2016-0004

Cerbin, W., & Kopp, B. (2006). Lesson Study as a model for building pedagogical knowledge and improving teaching. *International Journal of Teaching and Learning in Higher Education*, *18*(3), 250–257. https://www.isetl.org/ijtlhe/pdf/ijtlhe110.pdf

*Celik, A. O., & Guzel, E. B. (2020). How to improve a mathematics teacher's ways of triggering and considering divergent thoughts through Lesson Study. *International Electronic Journal of Mathematics Education*, *15*(3), 1–14. https://doi.org/10.29333/iejme/8461

*Chang, C.-H., Pascua, L., & Ess, F. (2018). Closing the "Hole in the Sky": The use of refutation-oriented instruction to correct students' climate change misconceptions. *Journal of Geography*, *117*(1), 3–16. https://doi.org/10.1080/00221341.2017.1287768

Cheung, W. M., & 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*, *3*(2), 137–149. https://doi.org/10.1108/IJLLS-05-2013-0024

*Chin Mon, C., Dali, M. H., & Chap Sam, L. (2016). Implementation of Lesson Study as an Innovative Professional Development Model Among Malaysian School Teachers. *Malaysian Journal of Learning and Instruction*, *13*(1), 83–111. https://doi.org/10.32890/mjli2016.13.1.5

*Chua, V. C. G. (2019). Going the distance: A lesson study on deriving the distance formula.

International Journal for Lesson and Learning Studies, *8*(2), 149–159. https://doi.org/10.1108/IJLLS-08-2018-0052

Clarke, D., & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education*, *18*(8), 947–967. https://doi.org/10.1016/S0742-051X(02)00053-7

*Coenders, F., & Verhoef, N. (2019). Lesson Study: Professional development (PD) for beginning and experienced teachers. *Professional Development in Education*, *45*(2), 217–230. https://doi.org/10.1080/19415257.2018.1430050

*Collet, V. S. (2017). Lesson Study in a Turnaround School: Local knowledge as a pressure-balanced valve for improved instruction. *Teachers College Record*, *119*(6). https://doi.org/10.1177/016146811711900605

*Collet, V. S., & Greiner, A. C. (2020). Revisioning grammar instruction through Collaborative Lesson Study: A new apprenticeship of observation. *Literacy Research and Instruction*, *59*(2), 95–120. https://doi.org/10.1080/19388071.2019.1709927

*Craney, C. L., Lau, T., Nelson, W. T., Ghomeshi, A., Rust, J. M., de Groot, R. M., Dennis, D. P., Robertson, E., & Kissel, J. (2020). Collaborative middle school science outreach project using the Japanese Lesson Study model. *Journal of Chemical Education*, *97*(5), 1256–1265. https://doi.org/10.1021/acs.jchemed.9b00780

*Csida, S., & Mewald, C. (2016). PrimarWebQuest in foreign language education. *International Journal for Lesson and Learning Studies*, *5*(1), 45–59. https://doi.org/10.1108/IJLLS-09-2015-0029

DeLuca, C., Shulha, J., Luhanga, U., Shulha, L. M., Christou, T. M., & Klinger, D. A. (2015). Collaborative inquiry as a professional learning structure for educators: A scoping review. *Professional Development in Education*, *41*(4), 640–670. https://doi.org/10.1080/19415257.2014.933120

Dewey, J. (1933). *How we think* (Original work published in 1910). Promotheus Books.

Dick, L. K., Appelgate, M. H., Gupta, D., & Soto, M. M. (2022). Continuous improvement Lesson Study: A model of MTE professional development. *Mathematics Teacher Educator*, *10*(2), 19. https://doi.org/10.5951/MTE.2020.0077

Dyke, M., Harding, A., & Lajeunesse, S. (2006, April). *Digital observation of teaching practice.* Annual Conference of the American Educational Research Association (AERA), San Francisco, CA. https://eprints.soton.ac.uk/24124/

Eisenhart, M., & DeHaan, R. L. (2005). Doctoral Preparation of Scientifically Based Education Researchers. *Educational Researcher*, *34*(4), 3–13. https://doi.org/10.3102/0013189X034004003

Elliott, J. (2012). Developing a science of teaching through lesson study. *International Journal for Lesson and Learning Studies*, *1*(2), 108–125. https://doi.org/10.1108/20468251211224163

*Estrella, S., Zakaryan, D., Olfos, R., & Espinoza, G. (2020). How teachers learn to maintain the cognitive demand of tasks through Lesson Study. *Journal of Mathematics Teacher Education*, *23*(3), 293–310. https://doi.org/10.1007/s10857-018-09423-y

*Færøyvik Karlsen, A. M. (2019). Investigating teacher learning in Lesson Study: The important link between reported observations and change of plans. *Professional Development in Education*, *48*(1), 53–69. https://doi.org/10.1080/19415257.2019.1685564

Fleetwood, S., & Hesketh, A. (2006). HRM-performance research: Under-theorized and lacking explanatory power. *The International Journal of Human Resource Management*, *17*(12), 1977–1993. https://doi.org/10.1080/09585190601041026

*Fox, A., & Poultney, V. (2020). Teacher professional learning through Lesson Study: Teachers' reflections. *International Journal for Lesson & Learning Studies*, *9*(4), 397–412. https://doi.org/10.1108/IJLLS-03-2020-0011

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal,* *38*(4), 915–945. https://doi.org/10.3102/00028312038004915

*Gee, D., & Whaley, J. (2016). Learning together: Practice-centred professional development to enhance mathematics instruction. *Mathematics Teacher Education and Development*, *18*(1), 87–99. https://doi.org/10.3316/aeipt.214245

*Gilissen, M. G. R., Knippels, M.-C. P. J., & van Joolingen, W. R. (2020). Bringing systems thinking into the classroom. *International Journal of Science Education*, *42*(8), 1253–1280. https://doi.org/10.1080/09500693.2020.1755741

*Goh, R., & Fang, Y. (2017). Improving English language teaching through Lesson sSudy: Case study of teacher learning in a Singapore primary school grade level team. *International Journal for Lesson and Learning Studies*, *6*(2), 135–150. https://doi.org/10.1108/IJLLS-11-2015-0037

Gough, D. (2007). Weight of evidence: A framework for the appraisal of the quality and relevance of evidence. *Research Papers in Education*, *22*(2), 213–228. https://doi.org/10.1080/02671520701296189

*Groves, S., Doig, B., Vale, C., & Widjaja, W. (2016). Critical factors in the adaptation and implementation of Japanese Lesson Study in the Australian context. *ZDM Mathematics Education*, *48*(4), 501–512. https://doi.org/10.1007/s11858-016-0786-8

Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching*, *8*(3), 381–391. https://doi.org/10.1080/135406002100000512

*Gutierez, S. B. (2015a). Collaborative professional learning through Lesson Study: Identifying the challenges of inquiry-based teaching. *Issues in Educational Research*, *25*(2), 118–134. https://doi.org/10.3316/informit.376688277371502

*Gutierez, S. B. (2015b). Teachers' reflective practice in Lesson Study: A tool for improving instructional practice. *Alberta Journal of Educational Research, 61*(3), 314–328. https://doi.org/10.11575/ajer.v61i3.56087

*Gutierez, S. B. (2016). Building a classroom-based professional learning community through lesson study: Insights from elementary school science teachers. *Professional Development in Education*, *42*(5), 801–817. https://doi.org/10.1080/19415257.2015.1119709

Hadfield, M., & Jopling, M. (2016). Problematizing Lesson Study and its impacts: Studying a highly contextualised approach to professional learning. *Teaching and Teacher Education*, *60*, 203–214. https://doi.org/10.1016/j.tate.2016.08.001

Hallinger, P. (2014). Reviewing Reviews of Research in Educational Leadership: An Empirical Assessment. *Educational Administration Quarterly*, *50*(4), 539–576. https://doi.org/10.1177/0013161X13506594

*Handayani, R. D., Wilujeng, I., Prasetyo, Z. K., & Triyanto. (2019). Building an indigenous learning community through lesson study: Challenges of secondary school science teachers. *International Journal of Science Education*, *41*(3), 281–296. https://doi.org/10.1080/09500693.2018.1548789

*Hapidin, Y., Hartati, S., & Nurani, Y. (2020). The continuous professional development for early childhood teachers through Lesson Study in implementing play based curriculum (Case study in Jakarta, Indonesia). *International Journal of Innovation*, *12*(10), 17–25.

Hardwicke, T. E., Wallach, J. D., Kidwell, M. C., Bendixen, T., Crüwell, S., & Ioannidis, J. P. A. (2020). An empirical assessment of transparency and reproducibility-related research

practices in the social sciences (2014–2017). *Royal Society Open Science*, *7*(2), 190806. https://doi.org/10.1098/rsos.190806

Hauge, K. (2021). The researcher's role: An intervention study using Lesson Study in Norway. In H. Senol (Ed.), *Pedagogy—Challenges, recent advances, new perspectives, and applications*. IntechOpen. https://doi.org/10.5772/intechopen.101100

*Howell, J. B., & Saye, J. W. (2016). Using Lesson Study to develop a shared professional teaching knowledge culture among 4th grade social studies teachers. *The Journal of Social Studies Research*, *40*(1), 25–37. https://doi.org/10.1016/j.jssr.2015.03.001

*Howell, J. B., & Saye, J. W. (2018). Integrating theory and practice: Factors shaping elementary teachers' interpretation of an inquiry model for teaching social studies. *The Journal of Social Studies Research*, *42*(2), 201–214. https://doi.org/10.1016/j.jssr.2017.04.003

*Huang, R., Gong, Z., & Han, X. (2016). Implementing mathematics teaching that promotes students' understanding through theory-driven Lesson Study. *ZDM: The International Journal on Mathematics Education*, *48*(4), 425–439. https://doi.org/10.1007/s11858-015-0743-y

*Huang, R., Kimmins, D., Winters, J., & Rushton, G. (2020). Does a technology-assisted Lesson Study approach enhance teacher learning while eliminating obstacles of traditional Lesson Study? *Contemporary Issues in Technology and Teacher Education, 20*(4), 618–659.

*Huang, R., Prince, K. M., & Barlow, A. T. (2017). Improving mathematics teaching as deliberate practice through Chinese Lesson Study. *The Mathematics Educator*, *26*(1), 32–55.

*Huang, R., Zhang, Q., Chang, Y., & Kimmins, D. (2019). Developing students' ability to solve word problems through learning trajectory-based and variation task-informed instruction. *ZDM*, *51*(1), 169–181. https://doi.org/10.1007/s11858-018-0983-8

Ishii, T. (2017). Historical overview of lesson study. In K. Tanaka, K. Nishioka, & T. Ishii (Eds.), *Curriculum, Instruction and Assessment in Japan: Beyond lesson study* (1st ed.). Routledge. https://doi.org/10.4324/9781315709116

Jacobs, V. R., Lamb, L. L., & Philipp, R. A. (2010). Professional noticing of children's mathematical thinking. *Journal for Research in Mathematics Education*, *41*(2), 169–202. https://doi.org/10.5951/jresematheduc.41.2.0169

*Jiang, H., Choy, B. H., & Lee, C. K.-E. (2020). Refining teaching expertise through analysing

students' work: A case of elementary mathematics teacher professional learning during lesson study in Singapore. *Professional Development in Education*, *46*(5), 731–750. https://doi.org/10.1080/19415257.2019.1634624

*Joubert, J., Callaghan, R., & Engelbrecht, J. (2020). Lesson Study in a Blended Approach to Support Isolated Teachers in Teaching with Technology. *ZDM: The International Journal on Mathematics Education*, *52*(5), 907–925. https://doi.org/10.1007/s11858-020-01161-x

Kager, K., Kalinowski, E., Jurczok, A., & Vock, M. (2021). *A systematic review of methodological transparency in Lesson Study: How do we report on the observation and reflection stages*? Open Science Framework. https://doi.org/10.17605/OSF.IO/5NXGY.

Kager, K., Jurczok, A., Bolli, S., & Vock, M. (2022). "We were thinking too much like adults": Examining the development of teachers' critical and collaborative reflection in lesson study discussions. *Teaching and Teacher Education*, *113*(103683), 1–13. https://doi.org/10.1016/j.tate.2022.103683

*Kanellopoulou, E.-M., & Darra, M. (2018a). The planning of teaching in the context of Lesson Study: Research findings. *International Education Studies*, *11*(2), 67–82. https://doi.org/10.5539/ies.v11n2p67

*Kanellopoulou, E.-M., & Darra, M. (2018b). The implementation of the Lesson Study approach to secondary education in Greece: The case of the literature lesson. *International Journal of Learning, Teaching and Educational Research*, *17*(7), 94–105. https://doi.org/10.26803/ijlter.17.7.6

*Kanellopoulou, E.-M., & Darra, M. (2018c). Making use of the Lesson Study approach to secondary education in Greece—The contribution to the professional development of teachers. *International Education Studies*, *11*(12), 78–85. https://doi.org/10.5539/ies.v11n12p78

*Karabuğa, F., & Ilin, G. (2019). Practicing lesson study in a Turkish education context: Considering the challenges, suggestions and benefits from EFL teachers' perspectives. *International Journal for Lesson and Learning Studies*, *8*(1), 60–78. https://doi.org/10.1108/IJLLS-05-2018-0036

*Karlsen, A. M. F., & Helgevold, N. (2019). Lesson Study: Analytic stance and depth of noticing in post-lesson discussions. *International Journal for Lesson and Learning Studies*, *8*(4), 290–304. https://doi.org/10.1108/IJLLS-04-2019-0034

Karlsen, A. M. F., & Ohna, S. E. (2021). Pupils' voices in teachers' collaborative professional

learning in Lesson Study. *International Journal of Educational Research*, *110*, 101877. https://doi.org/10.1016/j.ijer.2021.101877

*Khokhotva, O. (2018). Lesson Study in Kazakhstan: Case study of benefits and barriers for teachers. *International Journal for Lesson and Learning Studies*, *7*(4), 250–262. https://doi.org/10.1108/IJLLS-04-2018-0021

*Khokhotva, O., & Elexpuru Albizuri, I. (2019). Student voice in Lesson Study as a space for EFL teachers' learning: A case study in Kazakhstan. *International Journal for Lesson and Learning Studies*, *9*(2), 153–166. https://doi.org/10.1108/IJLLS-06-2019-0054

Kim, J. (2021). Through foreign eyes: A critical understanding of Lesson Study-based teacher education in Japan. In J. Kim, N. Yoshida, S. Iwata, & H. Kawaguchi (Eds.), *Lesson Study-based Teacher Education* (pp. 9–28). Routledge.

*Kıncal, R. Y., Ozan, C., & İleritürk, D. (2019). Increasing students' English language learning levels via Lesson Study. *English Language Teaching*, *12*(9), 88–94. https://doi.org/10.5539/elt.v12n9p88

*Klammer, S., & Hanfstingl, B. (2019). A first time Lesson Study that turned into a learning study. *International Journal for Lesson and Learning Studies*, *8*(4), 305–319. https://doi.org/10.1108/IJLLS-04-2019-0029

*Klefbeck, K. (2020). Lesson Study for students with intellectual disability. *International Journal for Lesson & Learning Studies*, *9*(3), 245–259. https://doi.org/10.1108/IJLLS-12-2019-0082

*Kohlmeier, J., Howell, J., Saye, J., McCormick, T., Shannon, D., Jones, C., & Brush, T. (2020). Investigating teacher adoption of authentic pedagogy through Lesson Study. *Theory & Research in Social Education*, *48*(4), 492–528. https://doi.org/10.1080/00933104.2020.1751761

*Koutsouris, G., Norwich, B., Fujita, T., Ralph, T., Adlam, A., & Milton, F. (2017). Piloting a dispersed and inter-professional Lesson Study using technology to link team members at a distance. *Technology, Pedagogy and Education*, *26*(5), 587–599. https://doi.org/10.1080/1475939X.2017.1364290

*Kunseeda, P., Inprasitha, M., & Changsri, N. (2018). Developing students' writing skills in mathematical problem-solving classroom using lesson study and open approach. *Asia-Pacific Journal of Science and Technology*, *23*(02), 1–8. https://doi.org/10.14456/kkurj.2018.3

Larssen, D. L. S., Cajkler, W., Mosvold, R., Bjuland, R., Helgevold, N., Fauskanger, J., Wood, P., Baldry, F., Jakobsen, A., Bugge, H. E., Næsheim-Bjørkvik, G., & Norton, J. (2018). A literature review of Lesson Study in initial teacher education: Perspectives about learning and observation. *International Journal for Lesson and Learning Studies*, *7*(1), 8–22. https://doi.org/10.1108/IJLLS-06-2017-0030

*Lawrence, M., Yang, L.-L., Briggs, M., Hession, A., Koussa, A., & Wagoner, L. (2016). Breathing life into engineering: A Lesson Study life science lesson. *Science Activities: Classroom Projects and Curriculum Ideas*, *53*(4), 137–146. https://doi.org/10.1080/00368121.2016.1211079

*Lee, L. H. J., & Tan, S. C. (2020). Teacher learning in Lesson Study: Affordances, disturbances, contradictions, and implications. *Teaching and Teacher Education*, *89*, 102986. https://doi.org/10.1016/j.tate.2019.102986

*Lee, V., & Madden, M. (2019). "We're in this together": Principals and teachers as partners and learners in Lesson Study. *NASSP Bulletin*, *103*(1), 51–64. https://doi.org/10.1177/0192636519826717

*Leifler, E. (2020). Teachers' capacity to create inclusive learning environments. *International Journal for Lesson & Learning Studies*, *9*(3), 221–244. https://doi.org/10.1108/IJLLS-01-2020-0003

*Lelinge, B., & Svensson, C. (2020). Teachers' awareness and understanding of students' content knowledge of geometric shapes. *Problems of Education in the 21st Century*, *78*(5), 777–798. https://doi.org/10.33225/pec/20.78.777

*Leong, S. S. M., Said, H. M., Shahrill, M., & Perera, J. S. H. Q. (2016). Using Lesson Study to enhance meaningful understanding on the topic of pressure. *International Journal of Environmental and Science Education*, *11*(15), 8425–8437.

Lewis, C. (2009). What is the nature of knowledge development in Lesson Study? *Educational Action Research*, *17*(1), 95–110. https://doi.org/10.1080/09650790802667477

Lewis, C. (2015). What Is Improvement Science? Do We Need It in Education? *Educational Researcher*, *44*(1), 54–61. https://doi.org/10.3102/0013189X15570388

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

https://doi.org/10.1007/978-3-319-07716-1_7

Lewis, C., Friedkin, S., Emerson, K., Henn, L., & Goldsmith, L. (2019). How does Lesson Study work? Toward a theory of Lesson Study process and impact. In R. Huang, A. Takahashi, & J. P. da Ponte (Eds.), *Theory and Practice of Lesson Study in Mathematics* (pp. 13–37). Springer International Publishing. https://doi.org/10.1007/978-3-030-04031-4_2

Lewis, C., & Perry, R. (2017). Lesson Study to scale up research-based knowledge: A randomized, controlled trial of fractions learning. *Journal for Research in Mathematics Education*, *48*(3), 261–299. https://doi.org/10.5951/jresematheduc.48.3.0261

Lewis, C., Perry, R., & Murata, A. (2006). How should research contribute to instructional improvement? The case of Lesson Study. *Educational Researcher*, *35*(3), 3–14. https://doi.org/10.3102/0013189X035003003 2006; 35

Lewis, J. (2016). Learning to lead, leading to learn: How facilitators learn to lead Lesson Study. *ZDM: The International Journal on Mathematics Education*, *48*(4), 527–540. https://doi.org/10.1007/s11858-015-0753-9

*Lewis, J., & Potts, P. (2019). Collaborative practitioner research: A sase of Lesson Study in a small jewish school. *Journal of Jewish Education*, *85*(2), 209–220. https://doi.org/10.1080/15244113.2019.1599609

*Lim, C. S., Kor, L. K., & Chia, H. M. (2016). Revitalising mathematics classroom teaching through Lesson Study (LS): A Malaysian case study. *ZDM Mathematics Education*, *48*(4), 485–499. https://doi.org/10.1007/s11858-016-0779-7

Limon, M. (2001). On the cognitive conflict as an instructional strategy for conceptual change: A critical appraisal. *Learning and Instruction*, *11*, 357–380. https://doi.org/10.1016/S0959-4752(00)00037-2

Littleton, K., & Mercer, N. (2013). *Interthinking: Putting talk to work*. Routledge.

*Liu, S.-H. (2016). Development of lesson study approach in three rural elementary schools of Taiwan. *Asia-Pacific Social Science Review*, *16*(2), 101–111.

*Loc, N. P., Uyen, B. P., Tong, D. H., & Ngoi, H. T. (2020). A teaching process of fostering students' problem-solving skills: A case study of teaching the equation of a line. *Universal Journal of Educational Research*, *8*(5), 1741–1751. https://doi.org/10.13189/ujer.2020.080510

*Lomibao, L. S. (2016). Enhancing mathematics teachers' quality through Lesson Study. *SpringerPlus*, *2016*(1), 1590. https://doi.org/10.1186/s40064-016-3215-0

*Lucenario, J. L. S., Yangco, R. T., Punzalan, A. E., & Espinosa, A. A. (2016). Pedagogical content knowledge-guided Lesson Study: Effects on teacher competence and students' achievement in chemistry. *Education Research International*, *2016*, 1–9. https://doi.org/10.1155/2016/6068930

*Lundbäck, B., & Egerhag, H. (2020). Lesson Study as a bridge between two learning contexts. *International Journal for Lesson & Learning Studies*, *9*(3), 289–299. https://doi.org/10.1108/IJLLS-02-2020-0006

Makel, M. C., & Plucker, J. A. (2014). Facts are more important than novelty: Replication in the education sciences. *Educational Researcher*, *43*(6), 304–316. https://doi.org/10.3102/0013189X14545513

Mann, S., & Walsh, S. (2013). RP or 'RIP': A critical perspective on reflective practice. *Applied Linguistics Review*, *4*(2), 291–315. https://doi.org/10.1515/applirev-2013-0013

Marton, F. (2015). *Necessary Conditions of Learning.* Routledge.

*Ma'rufi, M. R., & Muhammad, I. (2019). Pedagogical Content Knowledge (PCK) among mathematics teachers on function materials through Lesson Study in junior high school. *International Journal of Innovation, Creativity and Change*, *9*(1), 135–156.

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, *22*(3), 276–282.

Mercer, N. (2000). *Words and minds: How we use language to think together.* Routledge.

Meyrick, J. (2006). What is good qualitative research?: A first step towards a comprehensive approach to judging rigor/quality. *Journal of Health Psychology*, *11*(5), 799–808. https://doi.org/10.1177/1359105306066643

*Midgette, A. J., Ilten-Gee, R., Powers, D. W., Murata, A., & Nucci, L. (2018). Using Lesson Study in teacher professional development for domain-based moral education. *Journal of Moral Education*, 1–21. https://doi.org/10.1080/03057240.2018.1445982

*Moghaddam, A., Sarkar Arani, M. R., & Kuno, H. (2015). A collaborative inquiry to promote pedagogical knowledge of mathematics in practice. *Issues in Educational Research*, *25*(2), 170–186.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & and the PRISMA Group. (2009). Reprint—Preferred reporting items for systematic reviews and meta-analyses: The PRISMA Statement. *Physical Therapy*, *89*(9), 873–880. https://doi.org/10.1093/ptj/89.9.873

Moravcsik, A. (2020). *Transparency in qualitative research*. SAGE Publications Ltd.

https://doi.org/10.4135/9781526421036

*Moss, J., Hawes, Z., Naqvi, S., & Caswell, B. (2015). Adapting Japanese Lesson Study to enhance the teaching and learning of geometry and spatial reasoning in early years classrooms: A case study. *ZDM: The International Journal on Mathematics Education*, *47*(3), 377–390. https://doi.org/10.1007/s11858-015-0679-2

Myers, J. (2012). The effects of Lesson Study on classroom observations and perceptions of lesson effectiveness. *Journal of Effective Teaching*, *12*(3), 94–104.

Mynott, J. P. (2019). Lesson study outcomes: A theoretical model. *International Journal for Lesson and Learning Studies*, *8*(2), 117–134. https://doi.org/10.1108/IJLLS-08-2018-0057

*Ni Shuilleabhain, A., & Seery, A. (2018). Enacting curriculum reform through Lesson Study: A case study of mathematics teacher learning. *Professional Development in Education*, *44*(2), 222–236. https://doi.org/10.1080/19415257.2017.1280521

*Norwich, B., Fujita, T., Adlam, A., Milton, F., & Edwards-Jones, A. (2018). Lesson Study: An inter-professional collaboration approach for educational psychologists to improve teaching and learning. *Educational Psychology in Practice*, *34*(4), 370–385. https://doi.org/10.1080/02667363.2018.1468733

*Norwich, B., Koutsouris, G., Fujita, T., Ralph, T., Adlam, A., & Milton, F. (2016). Exploring knowledge bridging and translation in lesson study using an inter-professional team. *International Journal for Lesson and Learning Studies*, *5*(3), 180–195. https://doi.org/10.1108/IJLLS-02-2016-0006

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science*, *7*(6), 615–631. https://doi.org/10.1177/1745691612459058

*Obara, S., & Bikai, N. (2019). Promoting math teacher active learning with the lesson study approach: A case study of in-service teachers' perspectives. *International Journal for Lesson and Learning Studies*, *8*(2), 135–148. https://doi.org/10.1108/IJLLS-11-2018-0088

O'Connor, C., & Joffe, H. (2020). Intercoder reliability in qualitative research: Debates and practical guidelines. *International Journal of Qualitative Methods*, *19*, 160940691989922. https://doi.org/10.1177/1609406919899220

*Ogegbo, A. A., Gaigher, E., & Salagaram, T. (2019). Benefits and challenges of Lesson Study: A case of teaching physical sciences in South Africa. *South African Journal of Education*, *39*(1), 1–9. https://doi.org/10.15700/saje.v39n1a1680

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—A web and mobile app for systematic reviews. *Systematic Reviews*, *5*(1), 210. https://doi.org/10.1186/s13643-016-0384-4

*Özaltun Çelik, A., & Bukova Güzel, E. (2018). Describing Lesson Study designed for improvement of mathematics teachers' knowledge of student thinking. *International Journal for Mathematics Teaching and Learning*, *19*(2), 176–204.

*Özdemir, S. M. (2019). Implementation of the Lesson Study as a tool to improve students' learning and professional development of teachers. *Participatory Educational Research*, *6*(1), 36–53. https://doi.org/10.17275/per.19.4.6.1

*Pang, J. (2016). Improving mathematics instruction and supporting teacher learning in Korea through Lesson Study using five practices. *ZDM: The International Journal on Mathematics Education*, *48*(4), 471–483. https://doi.org/10.1007/s11858-016-0768-x

Parks, A. N. (2008). Messy learning: Preservice teachers' lesson-study conversations about mathematics and students. *Teaching and Teacher Education*, *24*(5), 1200–1216. https://doi.org/10.1016/j.tate.2007.04.003

*Pehlivan, F. C., & Guzel, E. B. (2020). Development of mathematics teachers' moves that support students' reasoning through Lesson Study. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*. https://doi.org/10.16949/turkbilmat.683535

*Pella, S. (2015). Pedagogical reasoning and action: Affordances of practice-based teacher professional development. *Teacher Education Quarterly*, *42*(3), 81–111.

Penuel, W. R., Fishman, B. J., Yamaguchi, R., & Gallagher, L. P. (2007). What makes professional development effective? Strategies that foster curriculum implementation. *American Educational Research Journal*, *44*(4), 921–958. https://doi.org/10.3102/0002831207308221

Piaget, J. (1985). *The Equilibrium of Cognitive Structures*. University of Chicago Press.

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, *66*(2), 211–227. https://doi.org/10.1002/sce.3730660207

*Rahim, S. S. A., Sulaiman, S., & Sulaiman, T. (2015). Teacher professional development through Lesson Study in secondary schools. *Advanced Science Letters*, *21*(7), 2360–2364. https://doi.org/10.1166/asl.2015.6276

R Core Team. (2013). *R: A language and environment for statistical computing.* R Foundation

for Statistical Computing. http://www.R-project.org/

Radovic, D., Black, L., Williams, J., & Salas, C. E. (2018). Towards conceptual coherence in the research on mathematics learner identity: A systematic review of the literature. *Educational Studies in Mathematics*, *99*(1), 21–42. https://doi.org/10.1007/s10649-018-9819-2

*Rochintaniawati, D., Riandi, R., Kestiniaty, J., Kindy, N., & Rukayadi, Y. (2019). The analysis of biology teachers' technological pedagogical content knowledge development in Lesson Study in West Java Indonesia. *Jurnal Pendidikan IPA Indonesia*, *8*(2). https://doi.org/10.15294/jpii.v8i2.19303

Rosenshine, B. (1994). The conduct and reporting of intervention studies. *Educational Psychology Review*, *6*(3), 245–254. https://doi.org/10.1007/BF02213186

*Rozimela, Y. (2020). Developing teachers' professionalism through school initiative-based Lesson Study. *European Journal of Educational Research*, *9*(4), 1513–1526. https://doi.org/10.12973/eu-jer.9.4.1513

*Sato, T. (2020). Japanese elementary teachers' professional development experiences in physical education lesson stud. *Physical Education and Sport Pedagogy*, *25*(2), 137-153. https://doi.org/10.1080/17408989.2019.1692808

Schipper, T., Goei, S. L., de Vries, S., & van Veen, K. (2018). Developing teachers' self-efficacy and adaptive teaching behaviour through lesson study. *International Journal of Educational Research*, *88*, 109–120. https://doi.org/10.1016/j.ijer.2018.01.011

Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. Basic Books.

Schön, D. A. (1995). *The Reflective Practitioner: How Professionals Think in Action*. Ashgate.

*Seino, T., & Foster, C. (2020). Analysis of the final comments provided by a knowledgeable other in lesson study. *Journal of Mathematics Teacher Education*, *24*(5), 507–528. https://doi.org/10.1007/s10857-020-09468-y

Seleznyov, S. (2018). Lesson study: An exploration of its translation beyond Japan. *International Journal for Lesson and Learning Studies*, *7*(3), 217–229. https://doi.org/10.1108/IJLLS-04-2018-0020

Seleznyov, S. (2019). Lesson study beyond Japan: Evaluating impact. *International Journal for Lesson and Learning Studies*, *8*(1), 2–18. https://doi.org/10.1108/IJLLS-09-2018-0061

Seleznyov, S., Goei, S. L., & Ehren, M. (2021). International policy borrowing and the case of Japanese Lesson Study: Culture and its impact on implementation and adaptation.

Professional Development in Education, 1–15. https://doi.org/10.1080/19415257.2021.1973069

*Shanmugam, S. K. S., Chinnappan, M., & Leong, K. E. (2020). Action research in examining the enquiry approach of Lesson Study in mathematics. *Pertanika J. Soc. Sci. & Hum.*, *28*(3), 1675–1693.

*Shi, L., & Cheng, E. C. K. (2020). Developing metacognitive teaching in Chinese language through conducting lesson study in Shanghai. *International Journal for Lesson & Learning Studies*, *10*(1), 75–88. https://doi.org/10.1108/IJLLS-09-2020-0065

*Sjunnesson, H. (2020). Initializing phase of lesson study: Communication a special didactic tool in mathematics. *International Journal for Lesson & Learning Studies*, *9*(3), 261–275. https://doi.org/10.1108/IJLLS-02-2020-0007

*Skott, C. K., & Møller, H. (2017). The individual teacher in lesson study collaboration. *International Journal for Lesson and Learning Studies*, *6*(3), 216–232. https://doi.org/10.1108/IJLLS-10-2016-0041

*Skott, C. K., & Møller, H. (2020). Adaptation of lesson study in a Danish context: Displacements of teachers' work and power relations. *Teaching and Teacher Education*, *87*, 102945. https://doi.org/10.1016/j.tate.2019.102945

Sherin, M. G., & Han, S. Y. (2004). Teacher learning in the context of a video club. *Teaching and Teacher Education*, *20*(2), 163–183. https://doi.org/10.1016/j.tate.2003.08.001

Stepanek, J., Appel, G., Leong, M., Turner Mangan, M., & Mitchell, M. (2007). *Leading Lesson Study: A practical guide for teachers and facilitators.* Corwin Press.

Stigler, J. W., & Hiebert, J. (2016). Lesson Study, improvement, and the importing of cultural routines. *ZDM: The International Journal on Mathematics Education*, *48*(4), 581–587. https://doi.org/10.1007/s11858-016-0787-7

*Suh, J., & Seshaiyer, P. (2015). Examining teachers' understanding of the mathematical learning progression through vertical articulation during Lesson Study. *Journal of Mathematics Teacher Education*, *18*(3), 207–229. https://doi.org/10.1007/s10857-014-9282-7

*Suhirman, L. (2018). Lesson study-based instruction for enhancing EFL teacher's pedagogical competence. *Asian EFL Journal*, *20*(5), 154–174.

*Suhirman, L. (2019). Building professional development for peer coaching and collaborating to plan research lesson: Case of Japanese Lesson Study approach. *The Asian EFL Journal*,

21(2.4), 9–32.

*Suhirman, L. (2020). Lesson Study—A Japanese instructional approach: Empowering student – centered Llearning. *Asian EFL Journal*, *27*(3.3), 209–229.

Takahashi, A. (2014). The role of the knowledgeable other in Lesson Study: Examining the final comments of experienced Lesson Study practitioners. *Mathematics Teacher Education and Development*, *16*(1), 4–21. https://doi.org/10.3316/aeipt.205651

*Tan, R. M. (2015). Improving the use of physical manipulatives in teaching science concepts through lesson study. *International Journal for Lesson and Learning Studies*, *4*(4), 328–342. https://doi.org/10.1108/IJLLS-01-2015-0001

*Thinwiangthong, S., Eddy, C. M., & Inprasitha, M. (2020). Mathematics teachers' abilities in developing formative assessment after the introduction of Lesson Study and Open Approach innovations. *Malaysian Journal of Learning & Instruction*, *17*(1), 101–132.

*Tsukui, A., & Saito, E. (2018). Stroll into students' learning: Acts to unload teachers' values through the practices of lesson study for learning community in Vietnam. *Improving Schools*, *21*(2), 173–186. https://doi.org/10.1177/1365480217717530

Ustuk, Ö., & De Costa, P. I. (2021). Reflection as meta-action: Lesson Study and EFL teacher professional development. *TESOL Journal*, *12*(1). https://doi.org/10.1002/tesj.531

*Vale, C., Widjaja, W., Doig, B., & Groves, S. (2019). Anticipating students' reasoning and planning prompts in structured problem-solving lessons. *Mathematics Education Research Journal*, *31*(1), 1–25. https://doi.org/10.1007/s13394-018-0239-5

Van Dijk, W., Schatschneider, C., & Hart, S. H. (2021). Open science in education sciences. *Journal of Learning Disabilities*, *54*(2), 139–152. https://doi.org/10.1177/0022219420945267

Van Es, E. (2011). A framework for learning to notice student thinking. In M. G. Sherin, V. R. Jacobs, & R. A. Philipp (Eds.), *Mathematics teacher noticing. Seeing through teachers' eyes* (pp. 164–181). Routledge.

VERBI Software. (2019). *MaxQda 2020*. VERBI Software. https://www.maxqda.com

*Verhoef, N. C., Coenders, F., Pieters, J. M., van Smaalen, D., & Tall, D. O. (2015). Professional development through lesson study: Teaching the derivative using GeoGebra. *Professional Development in Education*, *41*(1), 109–126. https://doi.org/10.1080/19415257.2014.886285

*Wang, H., Gibson, C., & Zander, U. (2020). Editors' comments: Is research on corporate social

responsibility undertheorized? *Academy of Management Review*, *45*(1), 1–6. https://doi.org/10.5465/amr.2019.0450

*Warwick, P., Vrikki, M., Færøyvik Karlsen, A. M., Dudley, P., & Vermunt, J. D. (2019). The role of pupil voice as a trigger for teacher learning in Lesson Study professional groups. *Cambridge Journal of Education*, *49*(4), 435–455. https://doi.org/10.1080/0305764X.2018.1556606

*Warwick, P., Vrikki, M., Vermunt, J. D., Mercer, N., & van Halem, N. (2016). Connecting observations of student and teacher learning: An examination of dialogic processes in Lesson Study discussions in mathematics. *ZDM*, *48*(4), 555–569. https://doi.org/10.1007/s11858-015-0750-z

*Watanabe, T., Takahashi, A., & Barham, A. I. (2019). *Implementing school-wide collaborative lesson research in Quatar*. *17*(2), 47–70. http://hdl.handle.net/10576/14352

*Widjaja, W., Vale, C., Groves, S., & Doig, B. (2017). Teachers' professional growth through engagement with lesson study. *Journal of Mathematics Teacher Education*, *20*(4), 357–383. https://doi.org/10.1007/s10857-015-9341-8

Wiggins, B. J., & Christopherson, C. D. (2019). The replication crisis in psychology: An overview for theoretical and philosophical psychology. *Journal of Theoretical and Philosophical Psychology*, *39*(4), 202–217. https://doi.org/10.1037/teo0000137

Willems, I., & Van den Bossche, P. (2019). Lesson Study effectiveness for teachers' professional learning: A best evidence synthesis. *International Journal for Lesson and Learning Studies*, *8*(4), 257–271. https://doi.org/10.1108/IJLLS-04-2019-0031

*Williams, J., & Ryan, J. (2020). On the compatibility of dialogism and dialectics: The case of mathematics education and professional development. *Mind, Culture, and Activity*, *27*(1), 70–85. https://doi.org/10.1080/10749039.2019.1686026

Wolgemuth, J. R., Hicks, T., & Agosto, V. (2017). Unpacking assumptions in research synthesis: A critical construct synthesis approach. *Educational Researcher*, *46*(3), 131–139. https://doi.org/10.3102/0013189X17703946

*Won, N. (2017). Inner-city teachers' perceptions in a Lesson Study for critiquing mathematical reasoning. *Professional Educator*, *42*(1), 58–79.

*Wood, K., Jaidin, H., Jawawi, R., Perera, J. S. H. Q., Salleh, S., Shahrill, M., & Sithamparam, S. (2017). How and what teachers learn from collaborative professional development. *International Journal for Lesson and Learning Studies*, *6*(2), 151–168.

https://doi.org/10.1108/IJLLS-09-2016-0028

*Woranetsudathip, N., & Yuenyong, C. (2015). Enhancing grade 1 Thai students' learning about mathematical ideas on addition through Lesson Study and Open Approach. *Mediterranean Journal of Social Sciences*, *6*(2), 28–33. https://doi.org/10.5901/mjss.2015.v6n2s1p28

Xu, H., & Pedder, D. (2014). Lesson Study: An international review of the research. In P. Dudley (Ed.), *Lesson Study: Professional learning for our time* (pp. 29-59). https://doi.org/10.4324/9780203795538

*Yildiz, A., & Baltaci, S. (2017). Reflections from the Lesson Study for the development of techno-pedagogical competencies in teaching fractal geometry. *European Journal of Educational Research*, *6*(1), 41–50. https://doi.org/10.12973/eu-jer.6.1.41

*Yoke, W. W., & Phang, F. A. (2017). Implementation of Lesson Study among physics teachers: A preliminary study. *Man in India*, *97*(13), 265–274.

Yoshida, N., Matsuda, M., & Miyamoto, Y. (2021). The landscape of Lesson Study: A methodology for teachers' professional development and educational research. In J. Kim, N. Yoshida, S. Iwata, & H. Kawaguchi (Eds.), *Lesson study-based teacher education: The potential of the japanese approach in global settings* (pp. 29–50). Routledge.

*Zhang, H., Yuan, R., & Liao, W. (2019). EFL teacher development facilitated by Lesson Study: A Chinese perspective. *TESOL Quarterly*, *53*(2), 542–552. https://doi.org/10.1002/tesq.480

*Zhang, Y. (2015). Sustaining Lesson Study in schools with positive peer leadership: A case study in Hong Kong. *International Journal for Lesson and Learning Studies*, *4*(2), 140–154. https://doi.org/10.1108/IJLLS-07-2014-0018

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). Academic Press.

Supplementary Material

Note: The present material is presented as supplementary material instead of an appendix due to the requirements of the journal that this article has been submitted to.

Table S1

Coding Protocol Used for the Assessment of Transparency of the Included Studies

Coding Protocol

Rater/Date: *Name, XX.XX.XXXX*

Quick Critical Appraisal Checklist (color in if fulfilled)

Published 2015-2020	Research design	In-service teachers (kindergarten to secondary school)	Published in a peer-reviewed journal in English	Lesson Study (LS) or a close adaptation*

Additional Information

Article tag	*author(s), year*
Article name	*title of article*
LS adaptation	*term used to refer to Lesson Study (e.g., LS, Japanese LS, Chinese LS, etc.) or reference to a LS handbook*
Country	*where study was conducted*
School level	*kindergarten, elementary/primary school, middle school, high school, other*
Impact of LS	*Is an impact reported? What kind of impact is reported? (include page number)*
Label for reflection stage	*e.g., post-lesson discussion, colloquium, debrief session, no specific label, etc.*
Research method	*e.g., content analysis, design-based research, action research (only include terms used in the article)*

Three-Level Rubric for Assessment of Transparency

Level	**Criteria**	**No information included**	**Partial information included**	**Detailed information included**	**Notes** (include page number)
		0	**1**	**2**	
Data Transparency	**Means of Data Collection** (**how** teachers collected data)	The article includes no reference to the means of data collection (that is, how teachers observed students and in what way these observations were recorded).	The article refers to the means of data collection (e.g., teachers used protocols, video recordings, structured notes, etc.).	The article explains the rationale behind the means of data collection and/or provides further material that renders the means of data collection explicit (e.g., appendix, table or figure, references or links to other sources).	

	Focus of Observation (**what** teachers observed)	The article includes no reference to the focus of teachers' observation(s) during the research lesson (that is, what teachers observed). *Note: mentioning LS' focus on student learning in passing does not count, as the term 'student learning' is vague and can be interpreted in multiple ways.*	The article includes some references to the focus of teachers' observation(s) during the research lesson (e.g., teachers observed how students reacted to a task or solved a task collaboratively).	The article elaborates on the focus of teachers' observation(s) (that is, it describes or explains why this focus was chosen or how it was pursued, e.g., through predictions or through indicators formulated prior to the research lesson).	
	Scope of Observation (**whom** teachers observed)	The article includes no information concerning whom teachers observed during the research lesson.	The article mentions whom teachers observed during the research lesson (e.g., the entire class, a specific subset of students, case students).	The article elaborates on whom teachers observed during the research lesson (that is, it explains the rationale behind why teachers observed specific students or the entire class, and/or mentions how case students were chosen).	
Analytic Transparency	**Interpretive Process** (how teachers **analyzed** and **reflected** on the data)	The article includes no conceptualization of teachers' interpretive process (that is, the article neither defines relevant terms, such as "reflection", nor does it theorize how teachers might interpret data together).	The article includes some references to how teachers analyzed their observations by providing examples of or quotations from the interpretive process. The reflection process is, however, neither defined nor explicitly explained.	The article conceptualizes teachers' interpretive process (e.g., by defining relevant terms, embedding them in a theoretical framework and explaining how this framework scaffolds teachers' meaning-making process(es)).***	
	Procedure/ Structure of Reflection Stage	The article includes no information on how the reflection stage was structured (that is, the steps or activities it comprised).	The article includes some information on how the reflection stage was structured (that is, some of its' steps or activities are mentioned).	The article elaborates on how the reflection stage was structured (that is, it includes a chronological and replicable description of the steps or activities involved).	

Production Transparency	**Role of Outside Expertise**	The article includes no information on whether or not external experts were present during the LS process.	The article states whether or not external experts were present (e.g., whether the author(s) acted as a "knowledgeable other(s)", or whether external partners facilitated the process).	The article elaborates on the role played by external experts (that is, it explains who they were, their roles, their participation or (likely) impact on the process).	
	Documentation of Reflection Stage	The article includes no information on whether or not teachers or external experts documented the reflection process (e.g., whether agreed-upon intentions were recorded in writing, whether teachers kept written reflections).	The article refers to the fact that teachers or external experts documented the reflection process in some way (e.g., observations and written reflections from the research lesson).	The article elaborates on who documented the reflection process in which way (e.g., observations and ideas were recorded in detail on charts, in order to be referred to during the next LS cycle).	
	Setting of the Reflection Stage	The article includes no information on the setting of the reflection stage (e.g., duration, location, other aspects).	The article includes some information on the setting of the reflection stage (e.g., teachers reflected together for two hours).	The article elaborates on the setting of teachers' reflection and includes details that make the process explicit.	

Total score: /16

The design of this three-level rubric is based on Hallinger (2014). The three levels of transparency are based on Moravcsik (2020)**. For our purposes, we have adapted the three-level rubric and the levels of transparency to examine transparency concerning the observation and reflection stages in Lesson Study.

*** For the purpose of this article, we understand LS or a close adaption of it as comprising the following elements (based on Seleznyov, 2018, p. 220-221):**

(1) The adaptation includes the following steps: teachers identify a focus, plan a lesson, teach the lesson and observe students, and discuss the observations (this may include the additional steps of re-teaching the lesson and then having a second discussion).

(2) The process is usually accompanied by some kind of outside expert(s) (e.g., project members, researchers, knowledgeable others, external facilitators, etc.).

(3) The article's authors explicitly refer to their PD model as LS or a particular form of LS (excluding Learning Study).

**** Definition of levels (adapted from Moravcsik, 2020)**

Data Transparency: Does the article communicate how teachers collected the observations during the research lesson?

Analytic Transparency: Does the article communicate the interpretive process by which teachers analyzed their observations and reflected together?

Production Transparency: Does the article communicate choices that framed the observation and reflection stages?

***** Summarized example from Brown et al. (2016, p. 9) that would be rated a 2:**

Brown et al. (2016) conceptualize the reflection stage in terms of high-quality learning conversations by referring to Stoll (2012) and deriving protocols and tools from Stoll which helped facilitate the teachers' conversations.

References:

Brown, C., Taylor, C., & Ponambalum, L. (2016). Using design-based research to improve the lesson study approach to professional development in Camden (London). *London Review of Education, 14*(2), 4–24. https://doi.org/10.18546/LRE.14.2.02

Hallinger, P. (2014). Reviewing Reviews of Research in Educational Leadership: An Empirical Assessment. Educational Administration Quarterly, 50(4), 539–576. https://doi.org/10.1177/0013161X13506594

Moravcsik, A. (2020). *Transparency in qualitative research.* SAGE Publications Ltd. https://doi.org/10.4135/9781526421036

Seleznyov, S. (2018). Lesson study: An exploration of its translation beyond Japan. *International Journal for Lesson and Learning Studies, 7*(3), 217–229. https://doi.org/10.1108/IJLLS-04-2018-0020

Stoll, L. (2012). Stimulating learning conversations. *Professional Development Today, 14* (4), 6–12.

Table S2

List of Keywords Used During the Coding Process

Level	Criteria	Keywords
Data Transparency	Means of Data Collection	protocol, template, handbook, video, note, writing, form, guide, checklist, rubric, artifact, student /pupil interview, pre-/post-test, student/pupil assessment, student/ pupil work, tool, score
	Focus of Observation	**theoretical framework for observation:** notic*, professional noticing, (professional) vision, observe*, (selective) attention student/pupil participation, student/pupil reaction, learning, indicator, predict*, anticipat*, student/pupil learning
	Scope of Observation	case student/pupil, whole/entire class, group of students/ pupils, subgroup
Analytic Transparency	Interpretive Process	reflect* **theoretical framework for reflection:** dialogue, dialogic space, teacher talk, language, reflection-in-action/ reflection-on-action, group conversation, cognitive conflict, inquiry
	Procedure/Structure of the Reflection Stage	e.g., first, second, third, then, finally, lastly, …
Production Transparency	Role of Outside Expertise	external, facilitat*, expert*, outside, researcher, guidance, prompts, scaffold, knowledgeable other, project leader/member, author
	Record-keeping	record*, note, protocol, template, presentation, journal, written reflection
	Setting of the Reflection Stage	minutes, min., hour, h., duration, lasted/lasting, length, period, teachers' lounge, staff room, empty (class)room, campus, immediately, same day

Table S3

List of Countries in Which the Studies Included in this Review Were Conducted (33 Countries Overall)

North America	***N* = 28**
Canada	4
US	24
Europe	***N* = 32**
Austria	2
Denmark	2
Greece	3
Ireland	1
Italy	1
Malta	1
Netherlands	3
Norway	2
Sweden	5
UK	12
Asia	***N* = 62**
Brunei	2
China	5
Hong Kong	1
Indonesia	10
Iran	1
Japan	2
Kazakhstan	2
Kingdom of Bahrain	1
Korea	1
Malaysia	5
Philippines	8
Qatar	1
Singapore	4
Taiwan	1

Thailand	3
Turkey	13
Vietnam	2
Oceania	***N* = 3**
Australia	*3*
Africa	***N* = 3**
South Africa	2
Eritrea	1
South America	***N* = 1**
Chile	*1*
Total	***N* = 129**

Table S4

List of School Levels in which the Studies Included in this Review Were Conducted

School level	
kindergarten/ prep school	3
kindergarten/ prep school and primary school	3
primary school	44
primary and secondary school	11
secondary school	61
primary, secondary and special school	1
center school	1
not specified	5
Total	***N* = 129**

Table S5

List of Self-Reported Research Designs Employed by the Studies Included in this Review

Research design	
Case Study (including single and collective case studies, and exploratory and narrative case studies)	60
Qualitative Design (not further specified)	48
Action Research	3
Design-based Research	4
Ethnographic Research	3
Personal Narrative	2
Phenomenological Approach	4
Qualitative Intervention Study	2
Quasi-experimental design	3
Mixed-Method design	5

Note. Some studies named more than one approach.

Table S6

List of Lesson Study Adaptations Named in the Studies Included in this Review

Lesson Study Adaptations	
LS	79
LS with adaptation (e.g., LS with distant technology, blended LS, technology-assisted LS, scaffolded LS, participatory LS, LS with computer-supported collaborative learning)	13
Japanese LS	12
LS with specific reference to the UK	11
Chinese LS	6
LS and Open Approach	3
Japanese LS with adaptation (e.g., Japanese-style scaffolded LS)	2
Lesson Study for Learning Community (LSLC)	2
Collaborative Lesson Research (CLR)	1
Total	***N* = 129**

Table S7

List of the 18 Studies Rated with 2 (i.e., Including Detailed Information) for the Category Means of Data Collection

No.	Author(s)	Year	Country	Means of Data Collection	page
1	Aydogan Yenmez et al.	2017b	Turkey	the LS group developed an assessment instrument based on their experiences from previous LS cycles, the final assessment tool is included in the article	p. 902
2	Baz	2020	Turkey	the LS group used a Group Performance Observation Form, included in the appendix	p. 73
3	Bruce et al.	2016	Canada	the LS group developed an observation guide, a sample observation chart and guiding questions are provided in the article	p. 549
4	Bütün	2019	Turkey	the author explains their rational behind taking observation notes in an extra column in the form of a lesson plan instead of videotaping the research lesson	p. 58
5	Collet & Greiner	2020	US	the LS process, including the observation, was guided by protocols and discussion templates from Collet (2019)	p. 102
6	Craney et al.	2020	US	the LS group developed a rubric to systematize classroom observations, access link for rubric is provided in the article	p. 1258
7	Færøyvik Karlsen	2019	Norway	the LS group developed two observation forms based on characteristics of exploratory talk (Mercer, 2004), both forms are included in the article	p. 4–5
8	Joubert et al.	2020	South Africa	the authors mention an observation schedule based on Bloom's taxonomy for learning, teaching, and assessing	p. 919
9	Kanellopoulou & Darra	2018a	Greece	the LS group used tools of observation and reflection adopted from Stepanek et al. (2007)	p. 79
10	Karlsen & Helgevold	2019	Norway	the LS groups developed their own observation forms, which are included in the article	p. 293
11	Khokhotva	2018	Kazakhstan	the author mentions the handbooks by Chichibu (2013) and Dudley (2014) as their main reference documents for the LS process, including the observation	p. 254
12	Khokhotva & Elexpuru Albizuri	2019	Kazakhstan	the LS group used observation templates based on Dudley (2014)	p. 157

13	Koutsouris et al.	2017	UK	the authors explain their decisions concerning where to position the video cameras for taping the research lesson	p. 591–593
14	Lawrence et al.	2016	US	the LS group observed the students and also used a scoring guide, this scoring guide was first used by students for self-assessment, and later by teachers, the scoring guide is included in article	p. 140
15	Lucenario et al.	2016	Philippines	the LS group used a LS Student Observation Form adapted from Weiland et al. (2010)	p. 5
16	Özdemir	2019	Turkey	the LS group used an observation form and took additional field notes, the author describes the aim of the observation form	p. 42
17	Pang	2016	Korea	the author states that the whole LS process was guided by the *Five Practices for Orchestrating Productive Mathematics Discussions* (Smith & Stein, 2011), the LS group used an analytic tool based on these practices for observation	p. 473
18	Sjunnesson	2020	Sweden	the LS group used an observation protocol based on Waldmann et al. (2015)	p. 167

Table S8

List of the Ten Studies that Theorized Teachers' Observations

No.	Author(s)	Year	Country	Theorization of the Observation Stage	page
1	Fox & Poultney	2020	UK	the authors connect the observation stage to the concept of "noticing" (Karlsen & Helgevold, 2019; Van Es, 2011) and the terms "focused noticing" and "extended noticing" (Karlsen & Helgevold, 2019)	p. 405–406
2	Huang et al.	2020	US	the authors shortly refer to the concepts of noticing and professional vision (Borko et al., 2014; Sherin & Han, 2004; Van Es et al., 2014) in the context of live observations being substituted with videos (technology-assisted LS)	p. 621
3	Karlsen & Helgevold	2019	Norway	the authors provide a detailed conceptualization of observation as "noticing" (Jacobs et al., 2010; Mason, 2002; Van Es, 2011) and connect noticing to the concept of interthinking (Littleton & Mercer, 2013)	e.g., p. 291–292

4	Klammer & Hanfstingl	2019	Austria	the authors describe how teachers created a graphical representation of the observed student learning; teachers' observations are connected to the inductive learning of students (a focus on how students learn and explore topics by themselves (Shaffer, 1989; Sik, 2015))	p. 309 – 311
5	Klefbeck	2020	Sweden	the author connects the observation stage and the visualization of students' learning to variation theory (e.g., Marton, 2015)	p. 247
6	Koutsouris et al.	2017	UK	the authors mention "tunnel vision" (Dyke et al., 2006) in connection with teachers' classroom observations and the LS groups' decision on recording modes (lessons were videotaped)	p. 595
7	Lee & Tan	2020	Singapore	the authors shortly report findings by Amador and Cater (2018) to define the professional noticing of LS groups	p. 4
8	Ni Shuilleabhain & Seery	2018	Ireland	the authors theorize the observation stage and the focus on student learning by referring to general theories on teacher learning (e.g., Penuel, 2014; Van Es & Sherin, 2008)	p. 224
9	Obara & Bikai	2019	US	the authors connect the observation stage to Garet et al.'s (2001) conceptualization of active learning	p. 137
10	Suh & Seshaiyer	2015	US	the authors connect the observation stage to "professional noticing" (Jacobs et al., 2010) as well as to "implicit and organic noticing" (Murata, 2011)	p. 211

Table S9

List of the 20 Studies that Theorized Teachers' Reflection Process

No.	Author(s)	Year	Country	Theorization of the Reflection Stage	page
1	Alwadi et al.	2020	Kingdom of Bahrain	the authors connect the reflection stage to practical thinking, they refer to Schön's (1995) knowledge-in-action and reflection-in-action	p. 335
2	Bozkurt & Yetkin-Özdemir	2018	Turkey	the authors conceptualize and define reflection based on e.g., Zimmermann (2000) and provide a detailed overview over reflection in teaching, with a focus on reflection activities	e.g., 379–381

3	Calleja & Formosa	2020	Malta	the authors define reflection as a cognitive process and connect reflection to the concept of cognitive conflict (e.g., Limon, 2001; Piaget, 1985; Posner et al., 1982)	p. 384–385
4	Cammarata & Haley	2018	Canada	the authors briefly refer to the role of reflection in teacher learning and cite Schön (1983)	p. 339
5	Collet & Greiner	2020	US	the authors define reflection as an important aspect of teacher learning and connect it to terms such as collaborative dialogue, critical discussions, and dialogic problem-solving process (Bakhtin, 1981; Horn & Little, 2010; Lave & Wenger, 1992; Wegerif, 2008)	p. 97–98, 113
6	Fox & Poultney	2020	UK	the authors speak of reflection within the framework of Illeris` (2011) and connect reflection to dialogue and cognitive dissonance (Mayrhofer, 2019; Mynott, 2019)	p. 399–400, 407
7	Goh & Fang	2017	Singapore	the authors briefly refer to Schön (1983) and his notion of the reflective practitioner	p. 138
8	Gutierez	2015b	Philippines	the author identifies reflection and critical reflection as important aspects of teachers' professional development (Darling-Hammond & Richardson, 2009; Larrivee, 2008) and also connect reflection to critical thinking (Conway, 2001)	p. 315–317
9	Gutierez	2016	Philippines	the author connects inquiry-based learning to reflective practice and further identifies reflection and the reflection stage in LS as a process of collaborative co-production of knowledge (Cornish & Jenkins, 2012)	e.g., p. 812
10	Jiang et al.	2020	Singapore	the authors do not define reflection itself but describe four patterns based on their findings which delineate teachers' inquiry process during the reflection stage in LS	p. 743–744
11	Khokhotva & Elexpuru Albizuri	2019	Kazakhstan	the authors conceptualize teacher learning based on transformative theory (Mezirow, 2000) and connect reflection to reflective dialogue (Jarvis, 2004)	p.155–156
12	Lim et al.	2016	Malaysia	the authors conceptualize reflection in the wider context of teacher change (Clarke & Hollingsworth, 2002) and provide a definition for working reflection	p. 487
13	Lomibao	2016	Philippines	the author anchors their study in Schön's (1983) conceptualization of reflection and links the notions of reflection-in-action and reflection-on-action to the observation and reflection stages in LS	p. 7
14	Pella	2015	US	the author anchors reflection in Shulman's (1987) notion of pedagogical reasoning and derives a definition for reflection	p. 84

15	Rozimela	2020	Indonesia	the author provides a detailed conceptualization and definition of reflection that is based a.o. on Cornford, 2002; Dewey, 1933; Farrell, 2007; and Rodgers, 2002.	p. 1515
16	Sato	2020	Japan	the author defines reflection based on Ryan and Deci (2000) and Mezirow (2000) as a critical element of teacher learning	p. 140
17	Seino & Foster	2020	Japan	the authors briefly mention the term “reflective practitioners” and refer the readers to Schön (1986) and Mason (2002)	p. 20
18	Suhirman	2019	Indonesia	the author provides a short paragraph on "reflective practice", but does not clearly anchor reflection in any existing theoretical framework	p. 16
19	Tsukui & Saito	2018	Vietnam	the authors base their understanding of reflection on Schön's (1983) notions of the reflective practitioner	p. 174
20	Won	2017	US	the authors speak of a “continued process of disequilibrium and reflection” in LS, which they explain by referring to Ball (1996) and the notions of critical discussions and collective inquiry (DuFour et al., 2006)	p. 2–3

Table S10

List of the Seven Studies that Provided an Alternative Theorization of the Reflection Stage

No.	Author(s)	Year	Country	Alternative Theorizations of the Reflection Stage	page
1	Bae et al.	2016	US	The authors connect the reflection stage to teacher talk and dialogue (Dudley, 2013)	p. 166
2	Brown et al.	2016	UK	the authors define "learning conversations" in professional learning communities based on Stoll (2012) and describe features that can make these learning conversations especially effective (learning conversations are not conceptualized in terms of reflection)	p. 9
3	Karlsen & Helgevold	2019	Norway	the authors anchor the reflection stage in the options of interthinking and modes of talk (Dudley, 2013; Littleton & Mercer, 2013; Mercer, 2000; Warwick et al., 2016)	p. 290–295
4	Lee & Tan	2020	Singapore	the authors talk about the reflection stage in terms of professional conversations (Nelson et al., 2010) and dialogic interactions (Suzuki, 2012; Warwick et al., 2016)	p. 2
5	Norwich et al.	2016	UK	the authors base their understanding of knowledge creation on the SECI model (Nonaka & Takeuchi, 1995), which is identified as enabling a structured, exploratory and reflective process	p. 191–192
6	Warwick et al.	2016	UK	the authors conceptualize a dialogic space in LS that enables teacher to interthink (Littleton & Mercer, 2013)	p. 557
7	Warwick et al.	2019	UK	the authors refer to two learning processes that have been observed in the reflection stage of LS: descriptive and interpretative learning processes (Vrikki et al., 2017)	p. 440

Additional References

These references are mentioned in the Supplementary Material but are not included in the main article's references.

Bakhtin, M. (1981). *The dialogic imagination.* (Translated and edited by M. Holquist and C. Emerson). University of Austin Press.

Ball, D. L. (1996). Teacher learning and the mathematics reforms: What we think we know and what we need to learn. *Phi Delta Kappan, 77*(7), 500–508.

Conway, P. F. (2001). Anticipatory reflection while learning to teach: From a temporally truncated to a temporally distributed model of reflection in teacher education. *Teaching and Teacher Education 17*, 89-106. https://doi.org/10.1016/S0742-051X(00)00040-8

Cornford, I. R. (2002). Reflective teaching: Empirical research findings and some implications for teacher education. *Journal of Vocational Education and Training, 54*(2), 219-235. https://doi.org/10.1080/13636820200200196

Cornish, L., & Jenkins, K.A. (2012). Encouraging teacher development through embedding reflective practice in assessment. Asia-Pacific journal of teacher education, 40(2), 159170. https://doi.org/10.1080/1359866X.2012.669825

Darling-Hammond, L., & Richardson, N. (2009). Teacher learning: What matters? *Educational Leadership, 66*(5), 46-53.

Dewey, J. (1933). *How we think.* Heath and Company.

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*, *34*, 107–121. https://doi.org/10.1016/j.tate.2013.04.006

DuFour, R., Eaker, R., & Many, T. (2006). *Learning by doing: A handbook for professional learning communities that work.* Solution Tree.

Farell, T. S. C. (2007). *Reflective language teaching from research to practice*. Continuum.

Horn, I. S., & 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), 181–217. https://doi.org/10.3102/0002831209345158

Illeris, K. (2011). *The Fundamentals of Workplace Learning: Understanding How People Learn in Working Life.* Routledge.

Jarvis, P. (2004), *Adult Education and Lifelong Learning: Theory and Practice*. Routledge.

Mason, J. (2002). *Researching your own practice: The discipline of noticing*. Routledge.

Mayrhofer, E. (2019). Lesson study and teachers' beliefs: How a Bourdieuian perspective could make a difference. *International Journal for Lesson and Learning Studies, 8*(1), 19-33. https://doi.org/10.1108/IJLLS-11-2018-0091

Mezirow, J. 2000. *Learning as Transformation: Critical Perspectives on a Theory in Progress*. Jossey-Bass.

Nelson, T. H., Deuel, A., Slavit, D., & Kennedy, A. (2010). Leading Deep Conversations in Collaborative Inquiry Groups. *Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *83*(5), 175–179. https://doi.org/ 10.1080/00098650903505498

Larrivee, B. (2008). Development of a tool to assess teachers' level of reflective practice. *Reflective Practice: International and Multidisciplinary Perspectives, 9*(3), 341-360, https://doi.org/10.1080/14623940802207451

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

Ryan, R. M., and E. L. Deci. 2000. Self-determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being. *American Psychologist, 55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Rodgers, C. (2002). Defining reflection: Another look at John Dewey and reflective thinking. *Teachers College Record, 104*(4), 842-866. https://doi.org/10.1111/1467-9620.00181

Schön, D. (1986). *Educating the reflective practitioner.* Jossey Bass.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of a new reform. *Harvard Educational Review, 57*(1), 1-22. https://doi.org/10.17763/haer.57.1.j463w79r56455411

Stoll, L. (2012). Stimulating learning conversations. *Professional Development Today*, *14*(4), 6–12.

Suzuki, Y. (2012). Teachers' professional discourse in a Japanese lesson study. *International Journal for Lesson and Learning Studies*, *1*(3), 216–231. https://doi.org/10.1108/20468251211256429

Wegerif, R. (2008). Dialogic or dialectic? The significance of ontological assumptions in research on educational dialogue. *British Educational Research Journal, 34*(3), 347–361. https://doi.org/10.1080/ 01411920701532228

Study 3

A Conceptual Model for Teachers' Continuous Professional Development through Lesson Study: Capturing Inputs, Processes, and Outcomes

Klara Kager, John Mynott & Miriam Vock

Kager, K., Mynott, J., & Vock, M. (2023). A Conceptual Model for Teachers' Continuous Professional Development through Lesson Study: Capturing Inputs, Processes, and Outcomes. *International Journal of Educational Research Open, 5*, 100272. https://doi.org/10.1016/j.ijedro.2023.100272

Abstract

Global interest in Lesson Study (LS), an iterative professional development model, is growing rapidly and has resulted in a rich body of findings that report mixed outcomes and impacts on teacher learning. In this conceptual paper, we argue that the field of LS currently lacks a conceptual model that can help tie these findings more closely to a common schematic and descriptive framework. Reviewing research on professional development, we derive the purpose of such a model and criteria that it should fulfil. We then examine current LS models, showing that several aspects, such as inputs, learning processes, LS' iterative character, and outcomes over time, are not sufficiently addressed. To fill these gaps, we draw on wider perspectives on teacher learning and organizational psychology and propose an updated model of LS. Lastly, we discuss concrete ways in which this model can be used in research and practice.

Key words: lesson study, professional development, conceptual model, learning outcomes, teacher learning

Introduction

Lesson Study (LS) is collaboration-based and teacher-driven approach to continuous professional development (PD). Over the span of several weeks, a group of teachers jointly investigate a problem of practice by studying the curriculum, planning a lesson, teaching and observing a live research lesson, and reflecting on their observations (Lewis et al., 2006). LS therefore includes several key characteristics of effective PD, that is, it addresses teachers' practice and real problems, focuses on students' learning, encourages collaboration and reflection, and is a sustainable and ongoing process (Borko et al., 2010). In the past three decades, LS has gained momentum across the globe and research reports that through LS teachers can, for instance, enhance their pedagogical and content knowledge (e.g., Coenders & Verhoef, 2019; Lewis et al., 2013), and increase their awareness for students' needs (Dudley, 2013).

There are, however, some tensions that surface repeatedly in the research literature. LS has been imported from its land of origin, Japan, to other education systems as a borrowed policy and adapted to fit diverse national and local contexts (Hadfield & Jopling, 2016; Seleznyov et al., 2021; Stigler & Hiebert, 2016). Not all LS adaptions are equally successful or produce similar outcomes (Adamson & Walker, 2011; Bjuland & Mosvold, 2015; Canonigo, 2016). In fact, how teachers learn within LS and its adaptations remains largely underconceptualized (Cheung & Wong, 2014; Elliott, 2012; Stigler & Hiebert, 2016) and crucial learning mechanisms, such as observation and reflection, are predominantly underdescribed in LS publications (Larssen et al., 2016; Kager et al., 2022). Sustaining LS practices over a long time period can prove challenging, and while several studies report assessments of the impact of LS (Dudley et al., 2019; Godfrey et al., 2019; Lewis & Perry, 2017; Schipper et al., 2020; Takahashi & McDougal, 2016), there seems to be no consensus on how to best evaluate LS outcomes (Cheung & Wong, 2014).

This means that we have accumulated a rich body of mostly descriptive and qualitative research on LS (Seleznyov, 2019; Xu & Pedder, 2014), yet it is difficult to systematically learn from its findings, as we lack both a shared conceptual framework of how local LS adaptations compare to one another as well as a language to talk about it. A conceptual model that systematically describes aspects that are potentially critical to continuous PD through LS and depicts long-term LS outcomes could establish such a common schematic framework for the field. The goal of such a model would be to connect diverse LS implementations, support the

development of a shared understanding of teachers' sustained learning through LS, and suggest avenues for future empirical research on LS.

The aim of this conceptual paper is to therefore develop a descriptive and theory-informed model of continuous PD through LS that systematically depicts its inputs, processes, and outcomes and can be used by both researchers and practitioners to assess short- and long-term impacts of LS. In a first step, we pinpoint what such a conceptual model should offer to the field. We then analyze the commonalities and differences of existing LS models and identify crucial issues that are currently insufficiently addressed, such as the means by which LS groups generate outcomes, as well as the emerging nature of these outcomes. To find ways to resolve these issues, we look beyond the field of LS and draw on influential models from research on professional development and organizational psychology. We then integrate these perspectives to propose an updated descriptive model that allows us to view continuous PD through LS not as a narrow and isolated event, but as a continuous, dynamic, and sustainable process that can and should be continuously evaluated and improved. Lastly, we discuss concrete ways in which this model can serve as a roadmap and a tool of analysis and evaluation for both researchers and practitioners.

The purpose of a conceptual model of Continuous Professional Development through Lesson Study

Darling-Hammond et al. (2017) emphasize that the goal of any professional development (PD) is to enhance teachers' knowledge and student learning. The primary concern when adopting a PD model, such as LS, is thus to test whether it can lead to these changes (Guskey, 2021). Testing a PD's effectiveness, however, presents several challenges. PDs are implemented in vastly different school contexts, which makes it almost impossible to replicate them without adaptations (Guskey, 2009). It is further inherently difficult to assess and quantify whether participation in a PD can lead to sustained changes in teacher's knowledge, since such changes, even if measurable, do not guarantee an immediate shift in teacher's daily practice (Korthagen, 2016) or an increase in student achievement (Guskey & Yoon, 2009). The evaluation of a PD and its outcomes is nevertheless crucial to ensure that the required resources are translated to a worthwhile outcome (Guskey, 2021; King, 2014).

Bryk (2015) argues that examining a PDs effectiveness might not be enough, especially for an iterative continuous improvement approach, such as LS (Lewis, 2015). Instead of asking

only whether an innovation works, it might make more sense to also ask which features need to be adapted or improved to make the innovation work for different agents under diverse conditions and over time (Bryk, 2015; Stigler & Hiebert, 2016). Even Guskey (2009, 2021), who advocates for rigorous assessments of PDs that yield replicable and comparative data, agrees that identifying and describing core elements that make PD effective, and ways in which they may be adapted, can be a productive way to circumvent the above-described challenges.

Along these lines, a group that perceives their LS work as ineffective does not need to immediately abandon the approach. They could, instead, assess factors that influenced their LS work, such as context conditions and how individual LS steps were implemented, and thereby pinpoint areas in which improvement or additional resources are needed. This formative and continuous evaluation could help ensure that LS can be a sustainable continuous PD model for diverse schools, instead of, in Lewis et al.'s (2006, p. 273) words, "a short-lived fad". This scenario presupposes, however, that the group has a clear understanding of LS and how to critically assess their achievements. In other words, they would need a model by which they can evaluate their outcomes and trace the steps by which they arrived there.

Concerning research, such a model could systematize how we describe and conceptualize LS, aid the theorization of LS (Stigler & Hiebert, 2016), as well as the development of a rich descriptive knowledgebase of LS (Lewis et al., 2006). Kitada (2022), who examined modifications of Japanese LS in the US context, argues that adaptations to LS are unavoidable and need to be taken into account holistically in research. As Kitada notes (2022), these adaptions are influenced at least in part by differences in ecological conditions and diverging teacher cultures, which impact the way we conceptualize PD in general and LS in particular. We therefore argue that a model that describes such conditions as well as teachers' learning processes and possible outcomes of LS could support the effort to methodically contextualize LS descriptions in research.

Such a model is currently missing in the LS literature and its development is challenged by the complexity of evaluating continuous PD. As Vanblaere and Devos (2021) note, assessing school improvement through continuous PD is difficult in general: the assessment needs to be long-term and rich in description, identify different developmental stages, and allow for comparisons of these stages in order to better understand what characterizes them. Davidoff et al. (2015) suggest that the use of a shared theory or conceptual framework can bring a

research field together by, first, systematizing features and their conditions that are crucial to an event, and, second, by ensuring that researchers are, in fact, investigating the same object of interest. A conceptual model in particular allows for the simplification of a complex event and provides a visual representation that ties research together (Jaakkola, 2020).

Based on the reviewed evidence on PD and models of assessment, we posit that a model of LS that could serve as a conceptual grid to various stakeholder groups would need to

- be applicable to different cultural contexts, LS adaptations, and subject areas,
- systematically describe the context factors that influence the implementation of LS, the LS steps and processes, and evolving short- and long-term outcomes,
- be useable for researchers to frame and explain their research, as well as to pinpoint areas of further research interest,
- and be useable for researchers and practitioners to conduct continuous and formative evaluation of LS cycles.

As a next step, we will review current models of LS and assess their suitability to address the above identified criteria.

Review of current models of Professional Development through Lesson Study

A survey of the literature indicates that the most frequently used LS model is circular and focuses exclusively on the LS core stages of study, plan, teach, and reflect, or variations thereof (e.g., Arani, 2006; Gutierez, 2016; Celik & Guzel, 2020; Chua, 2019; Dick et al., 2022; Dudley, 2013; Fujii, 2014; Isoda, 2015; Joubert et al., 2020; Lewis, 2009; Moss et al., 2015). These circular models provide a useful description of how LS steps are conceptualized, yet they largely leave contextual factors, specific learning processes, and learning outcomes in the dark.

Some models, summarized in Figure 1, extend beyond the circular illustration of LS' core stages. They usually aim to conceptualize how teachers learn through LS and frequently follow a linear structure that resembles an input-process-output model (I-P-O model). The I-P-O model (Hackman, 2012; Hackman & Morris, 1975) is traditionally considered a useful paradigm to conceptualize how group interaction processes are influenced by input factors and yield certain outputs.

According to Driskell et al. (2018), input refers to contextual conditions and participants' characteristics. This dimension is regarded as crucial in research on teacher learning, as factors

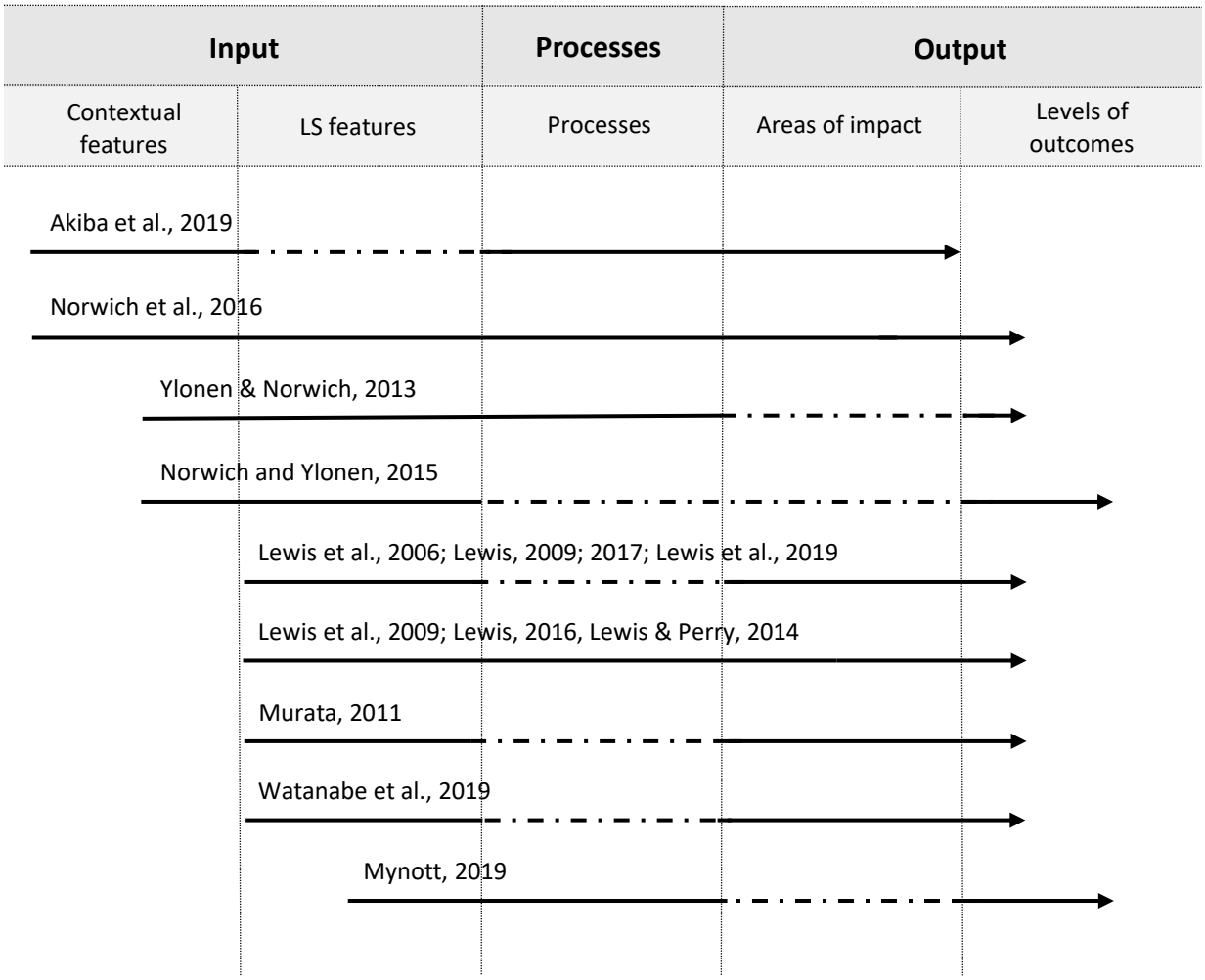
such as the school and classroom context, participants' knowledge and motivation, material quality, and resources have been shown to affect the outcomes of PDs (Borko, 2004; Darling-Hammond et al., 2017). Process links the dimensions of input and output by detailing how groups generate results (Driskell et al., 2018). This dimension is frequently described in terms of a "black box" (Cajkler et al., 2013; Hargreaves, 2005; Vrikki et al., 2017), due to the difficulty of analyzing learning processes. This challenge extends to the third dimension, output, which describes the results achieved by a group (Driskell et al., 2018) and is also commonly referred to as outcomes. The nature of these outcomes can be manifold and commonly include, for instance, participants' reactions and learning, the application of new learning, and students' learning outcomes (Darling-Hammond et al., 2017; Guskey, 2002). Stake and Schwandt (2006) further speak of a difference between the quality of results as measured (e.g., against a set standard) and as perceived by participants (e.g., gathered from participants' personal reflections).

Returning to Figure 1, we see that most models in the LS literature structured in line with the I-P-O model neglect one or more of these dimensions. First, the majority of models starts with structural features of LS, excluding any additional contextual aspects. Exceptions are, for example, Norwich and Ylonen (2015), who include "Lesson Study conditions and context" in their model, and Akiba et al. (2019), who specify three areas of input factors: duration, facilitator orientation, and material quality.

Similarly, the majority of models do not explicitly outline the means by which outcomes are achieved, leaving the dimension of process underconceptualized. An example is the influential model proposed by Lewis and colleagues (2006), which has been advanced in several subsequent publications. The model posits areas of "intervening changes", such as teachers' knowledge and commitment. The specific processes that induce these changes, however, remain largely unspecified. Two versions of the model that do address processes suggest that changes ensue as the group develops an identity and thinking becomes visible (Lewis et al., 2009), or through the collaborative study of materials (Lewis, 2016). These specifications, however, are not consistent and no longer included in the more recent version of the model (Lewis et al., 2019).

Figure 1

Analysis of the components and scope of existent Lesson Study models that follow the I-P-O structure



Notes. The dashed line indicates that this aspect is not included in a given model. The varying starting and ending point indicate to what extent a certain aspect is illustrated in a given model (e.g., most models describe one or two outcomes, while Norwich and Ylonen (2015) and Mynott (2019) define four to five outcomes).

One aspect largely missing from the models in Figure 1 is the iterative character of LS. In fact, only the model by Murata (2011) includes an arrow that (re)connects the dimensions of output and input, indicating that teachers' new insights inform future LS processes. Mathieu et al. (2019) note that temporal aspects are often overlooked in the traditional I-P-O model. Repeated cycles are, however, a crucial feature of LS (Seleznyov, 2018), and models that

neglect this aspect raise a host of questions. For example, it remains unclear in the models by Mynott (2019) and Norwich and Ylonen (2015) whether the proposed outcomes ensue after one or multiple LS cycles, whether they build on each other, or whether they represent different stages.

Next to the models listed in Figure 1, the LS literature offers some other models that circumvent the issue of iteration by either depicting multiple LS cycles or favoring a circular structure over the I-P-O structure. Examples of the former are the models by Dudley (2019) and Ylonen and Norwich (2013), which do not include all I-P-O dimensions and seem to be tailored towards local and research-specific versions of LS. Examples of the latter frequently draw on Clarke and Hollingsworth's (2002) interconnected model of teacher change (e.g., Bae et al., 2016; Da Ponte et al., 2022; Schipper et al., 2017; Widjaja et al., 2017). This model defines four domains (personal, external, practice, and consequence) and posits that professional learning is a non-linear but dynamic development embedded in these domains and driven by the processes of enactment of reflection (Clarke & Hollingsworth, 2002). The model's main focus lies on conceptualizing the process of professional learning, and as such, the model does not categorize the development of learning outcomes over time, or explicitly include the enhancement of student learning as a result of teachers' learning process. LS models that are based on Clarke and Hollingsworth's (2002) model consequently cannot offer strong explanatory power regarding different stages of outcomes and their development over time.

Figure 1 shows that existing LS models generally underconceptualize the dimension of learning outcomes. In most cases, outcomes are defined as areas of impact (such as teachers' content knowledge, or beliefs), and illustrated as a rather fixed sequence of changes (teachers' instructional improvement leads to improved student learning). This connects to the vivid debate on whether teacher change follows a specific linear structure (Desimone, 2009; Guskey, 2002) or should be conceptualized as an interconnected and more dynamic process (Clarke & Hollingsworth, 2002; Opfer & Pedder, 2011). It further demonstrates that the base models, on which we draw to inform LS models, come with limitations that will necessarily influence the scope of the LS model.

As Boylan et al. (2018) explain, there is a difference between general models of professional learning (Clarke & Hollingsworth, 2002; Desimone, 2009; Guskey, 2002) and models that classify specific professional learning outcomes (Guskey, 2000; Kennedy, 2005). Our review of

the literature indicates that current LS models tend to be adaptions of the former, rather than the latter. That is, they tend to define specific areas of change, such as “mathematics standards” (Watanabe et al., 2019, p. 51) or “[teachers’] self-efficacy” (Akiba et al., 2019, p. 354), and some add one or two linear outcomes (e.g., changes in the area of teachers’ knoweldge and beliefs will lead to instructional improvements, Lewis et al., 2009). As a consequence, these models cannot account for short-, mid- and long-term outcomes, run the risk of becoming too narrow for broad application, and tend to depict a linear view of professional learning nowadays considered as insufficient.

Our review of LS models is by no means exhaustive, but it demonstrates the challenges in modeling a dynamic process in a comprehensive way and indicates the need for the advancement of current approaches. To sum up, a number of conceptual models, aiming to describe varying parts and processes of LS, have been suggested, and each model marks an important contribution to our current understanding of LS. On the surface, many of these models follow the I-P-O structure, which corresponds to the traditional paradigm of evaluating collaborative processes (Driskell et al., 2018; Hackman, 2012). These models do not, however, share a common starting point (Figure 1) and tend to emphasize on some dimensions, while others remain underconceptualized. Additionally, some models focus on a LS adaptation specific to a piece of research or subject area (Da Ponte et al., 2022; Dudley et al., 2019; Norwich & Ylonen, 2013; Watanabe et al., 2019), making it difficult to translate it to other LS contexts. We also saw that LS models based on Clarke and Hollingsworth’s (2002) model tend to underdescribe emerging outcomes.

Current models of LS therefore fall short on all three aspects identified by Vanblaere and Devos (2021): they allow only for a limited assessment of long-term outcomes, and they neither describe different developmental stages of collaborative work, nor do they allow for the comparisons of these stages within and between schools. However, the synthesis of models demonstrates that the I-P-O structure is generally viewed as a suitable model for LS research, and several relevant inputs, processes, and areas of outcomes have already been identified.

At this point, it should be noted that the reviewed LS models were developed within their own specific contexts and provide a highly beneficial abstraction of LS for their purposes. They were not created with the explicit intention of offering a unifying conceptual model for PD through LS, or to meet the criteria we outlined in our introduction. We contend, however,

that the versatile and international field of LS has progressed to a point where it is possible, and also necessary, to develop a model of continuous PD through LS that incorporates the benefits of these existing models and attempts to capture all three dimensions.

Next, we will draw on research beyond the field of LS to identify suitable solutions to address these gaps. Specifically, we are seeking ways to represent the iterative structure of LS and coherently describe inputs, processes, and developing stages of outcomes.

Applying models of Professional Development to the field of Lesson Study

There are several influential models in and outside the field of education that can help advance our current set of knowledge on LS models. As already established, there is a difference between models that provide a general conceptualization of professional learning, and those that seek to classify professional learning outcomes (Boylan et al., 2018). It seems that a model of continuous PD through LS that comprehensively describes both the LS intervention and its impact would need to reconcile these two approaches.

In this section, we will therefore first consider research on teacher learning, drawing specifically on the work of Guskey (2000, 2002, 2021), which continuous to shape our discussion on the evaluation of PD. Given that LS is a process built on cooperation and that its outcomes are the effort of intense team work, we then take into account pertinent findings from research into team effectiveness and group work (e.g., Ilgen et al., 2005; Marks et al., 2001; Mathieu et al., 2019). The offer-and-use model for PD developed by Lipowsky and Rzejak (2015) has demonstrated that the perspective of organizational psychology can support the conceptualization of PD outcomes in the field of education. We therefore aim to bring together these perspectives from across disciplines in order to advance how we view teachers' continuous PD through LS.

Prior to his prominent model on teacher learning, Guskey (2000) formulated five hierarchical levels of outcomes as a way to systematically document and evaluate PDs. These levels include the teachers' reactions to, and their satisfaction with, a PD program, changes in their knowledge, changes in organizational support on a school-level, changes in teachers' daily practice, and finally change in students' assessments and grades. With these levels, Guskey (2000) proposes a likely sequence of how PDs can lead to immediate and long-term outcomes for several stakeholders. Each level also acts as a precondition for the next level. If

teachers are not satisfied with the PD program, for instance, it is unlikely that they will have capitalized on the provided learning opportunities or make changes to their practice.

Guskey's five levels still inform educational research today and have previously been used for the evaluation of LS outcomes. Seleznyov (2019), for instance, used an adapted form to analyze existing findings on LS impacts, showing that there is currently a dearth of studies that rigorously assess mid- and long-term outcomes. One reason for the predominant focus of LS research on Guskey's first level, the participants' reactions, might be that the individual LS cycle is too narrow of a time window to expect or measure changes of subsequent levels (Mynott, 2019). As Mynott (2019) argues, for changes to occur on the organizational or student level, teachers need to engage in LS over a longer period of time.

This emphasizes the gap we earlier identified in the literature: current models do not depict outcomes over time, which might be challenging our conceptualization of how these outcomes could look like, or be measured. In their extensive study on LS impact, Godfrey et al. (2019) showed that Guskey's five levels can serve as a useful heuristic to not only evaluate, but to plan and guide LS from the start. These examples suggest that Guskey's (2000) levels of outcomes could be a helpful schematic not only for individual research studies, but as part of a conceptual model that is shared within the research community.

While helpful in this regard, Guskey's (2000) levels focus exclusively on outcomes, omitting the dimensions of input and process, and also posit a linear sequence. The more recent model by Lipowsky (2014) and Lipowsky and Rzejak (2015) progresses Guskey's (2000) approach by embedding hierarchical outcome levels in an offer-and-use model. The offer-and-use model, similarly to the I-P-O model, provides a systematization of factors that have been shown to influence the effectiveness of a certain learning offer (i.e., a lesson, a workshop) and of the outcomes that the use of this learning offer can lead to. In the context of PD, these factors include the characteristics of the facilitator and the participants, the school context, the PD's structural aspects, and whether or not participants capitalized on the learning moments provided during the PD (Lipowsky & Rzejak, 2015). These interrelated aspects then lead, through a transfer process, to various outcome levels: participants' reaction and satisfaction, the enhancement of participant's knowledge and their instructions, and finally the development of students' performance.

Lipowsky and Rzejak (2015) therefore make to two crucial changes to Guskey's (2000) approach. First, they connect PD outcomes to the dimensions of input and process on a

conceptual level. Second, they indicate that outcome levels 2 (changes in teachers' knowledge and beliefs) and 3 (changes in teachers' practice) develop in parallel. This appears to be an effort to soften the implication that these outcomes evolve in a strictly linear fashion. Despite these changes, the model struggles to factor in the cyclical structure of teachers' collaborative continuous improvement and does not explain what a transfer process could entail. Given the omission of Guskey's (2000) outcome level 3 (organizational support and change), the model further neglects the impact teachers' continuous development might have on the organization they are embedded in, and also diminishes the organizations' role in providing the structures and systems necessary to uphold changes. Studies have shown repeatedly that administrative support and resources, such as time and space, can make or break a LS group's efforts (Godfrey et al., 2019; Groves et al., 2016; Lee & Tan, 2020; Lim et al., 2016).

As Boylan et al. (2018) note, models such as those by Clarke and Hollingsworth (2002) and Guskey (2002) - and as we argue also by Lipowsky and Rzejak (2015) - seem limited in their ability to explain the collaborative learning of teachers. At this point, we therefore turn to research on organizational groups and group effectiveness.

Similar to the field of education, research on group effectiveness makes frequent use of the I-P-O model, but has updated the model in order to better fit with the conceptualization of groups as complex, adaptive structures that evolve over time (Driskell et al., 2018; Mathieu et al., 2019). Specifically, there has been a shift in terminology, as the dimension of process tends to nowadays be referred to as mediating mechanisms (Ilgen et al., 2005). This shift was triggered by the argument that it is not only behavioral processes or acts that turn inputs into outputs, but also emergent cognitive and affective states (Marks et al., 2001). The term mediating mechanisms refers to both behavioral processes and emergent states that evolve as groups collaborate, such as group cohesion, trust, climate, and self-efficacy (Mathieu et al., 2019). A second shift in terminology responded to the challenge of conceptualizing dynamic developments along a linear and causal structure. Ilgen et al. (2005) explain that, in order to account for outputs as feedback loops that inform future inputs, an additional "I" (input) was added to the model, which resulted in the term IMOI-model. Further, the hyphens between letters were omitted to signal that "causal linkages may not be linear or additive, but rather nonlinear or conditional" (Ilgen et al., 2005, p. 520).

These small changes in terminology significantly expand the scope and usability of the model. They also help to better understand existent models of LS that have addressed the

dimension of mediating mechanisms. The versions of Lewis' model that include this dimension, for instance, appear to focus on cognitive and affective states, rather than processes. Lewis et al. (2009) note that intervening changes ensue as "community norms, tools, identity, and [teachers'] participation develop", while Lewis and Perry (2014) describe that teachers "assimilate and accommodate knowledge/ beliefs in response to materials, colleagues, students". In both cases, the model specifies dynamic properties of a group that emerge gradually. Other existent models concentrate on behavioral processes, such as the study of materials (Lewis, 2016; Norwich & Ylonen, 2013), observation and planning (Norwich & Ylonen, 2013), reflection (Bae et al., 2016; Da Ponte et al., 2022; Norwich et al., 2016), or enactment (Bae et al., 2016; Da Ponte et al., 2022; Norwich & Ylonen, 2013). The distinction between processes and emergent states therefore seems a viable solution to circumvent the limited ability of traditional PD models to account for collaborative and co-evolving aspects of learning (Boylan et al., 2018).

Turning to the dimension of outcomes, the IMOI-model offers two distinct advantages over the I-P-O model. First, the added "I" incorporates outcomes as future inputs in the very structure of the model. While this might not be the most elegant solution, and the visual structure of the model remains a linear line, it still signals the circularity of group processes and illustrates that each dimension, even input factors, develops and evolves over time.

Second, given that group effectiveness research is typically conducted in the context of industry or business organizations, outcomes tend to include participants' satisfaction and commitment, but also quality and efficiency of performance (Driskell et al., 2018; Mathieu & Gilson, 2012). Mathieu et al. (2019, p. 18) classify these outcomes into two distinct types. They speak of tangible outputs or products, which include productivity (quantity), efficiency (quantity relative to a set goal), and quality (value or worth). These types of outcomes need some period of time to develop and grow, and allow for a certain quantification of a group's output (Mathieu et al., 2019).

Next, Mathieu et al. speak of influences on the individual participants as well as on the collective group. These include changes in participant's attitudes, knowledge, or behavior. On a collective level, Mathieu et al. (2019) list cohesion and psychological safety, which refer to experiences shared by the group. Both cohesion and psychological safety could also be conceptualized as emergent states along the dimension of mediating mechanisms, which

emphasizes the temporal development of these categories and shows that the boarders between the IMOI dimensions are gradual and subject to definition (Mathieu et al., 2019).

Conceptualizing outcomes in terms of these two categories – tangible outcomes and influence on participants – makes it possible to evaluate LS from various perspectives. As Elliott (2019) points out, there is no straight-forward way to measuring the quality of LS outcomes and research, as quality-as-measured might differ from the quality-as-experienced by the LS participants. A LS model that makes the proposed distinction would allow to take both sides into account. Tangible outcomes, for instance, would include the number of LS cycles completed (productivity), the number of LS cycles completed in relation to a certain goal or project context (efficiency), and the quality of these cycles or value of produced materials and lesson plans (quality). Such a quantitative assessment might be valuable for a school or project group in order to assess and document progress, communicate their productivity to school boards or policy makers, and could be complemented by an analysis of outcomes in regard to the participants individual and collective reactions (i.e., participants' satisfaction with their LS work or changes in their attitudes, measured through e.g., surveys or pre- and post-tests).

In this section, we have drawn on several influential perspectives from outside the field of LS. We have discussed the importance of combining views of professional learning with classifications of learning outcomes, and we have explained how research on group effectiveness has addressed some of the gaps in current LS models. We will integrate these aspects in order to propose a conceptual model of continuous PD through LS.

A conceptual model of continuous Professional Development through Lesson Study

In the beginning of this paper, we derived that a shared model of LS should ideally be usable across cultural contexts and adaptations, provide a coherent description of input factors, processes, and various stages of outcomes, and be helpful for both researchers and practitioners to plan, implement, analyze, and evaluate LS. Based on the review and analysis of the wider literature, we now propose a model of LS that is an extension of current LS models and significantly advances the conceptualization of each of the I-P-O dimensions. We further update the structure of the model by using the IMOI framework, rather than the I-P-O framework, in order to include the circular structure of LS and differentiate between

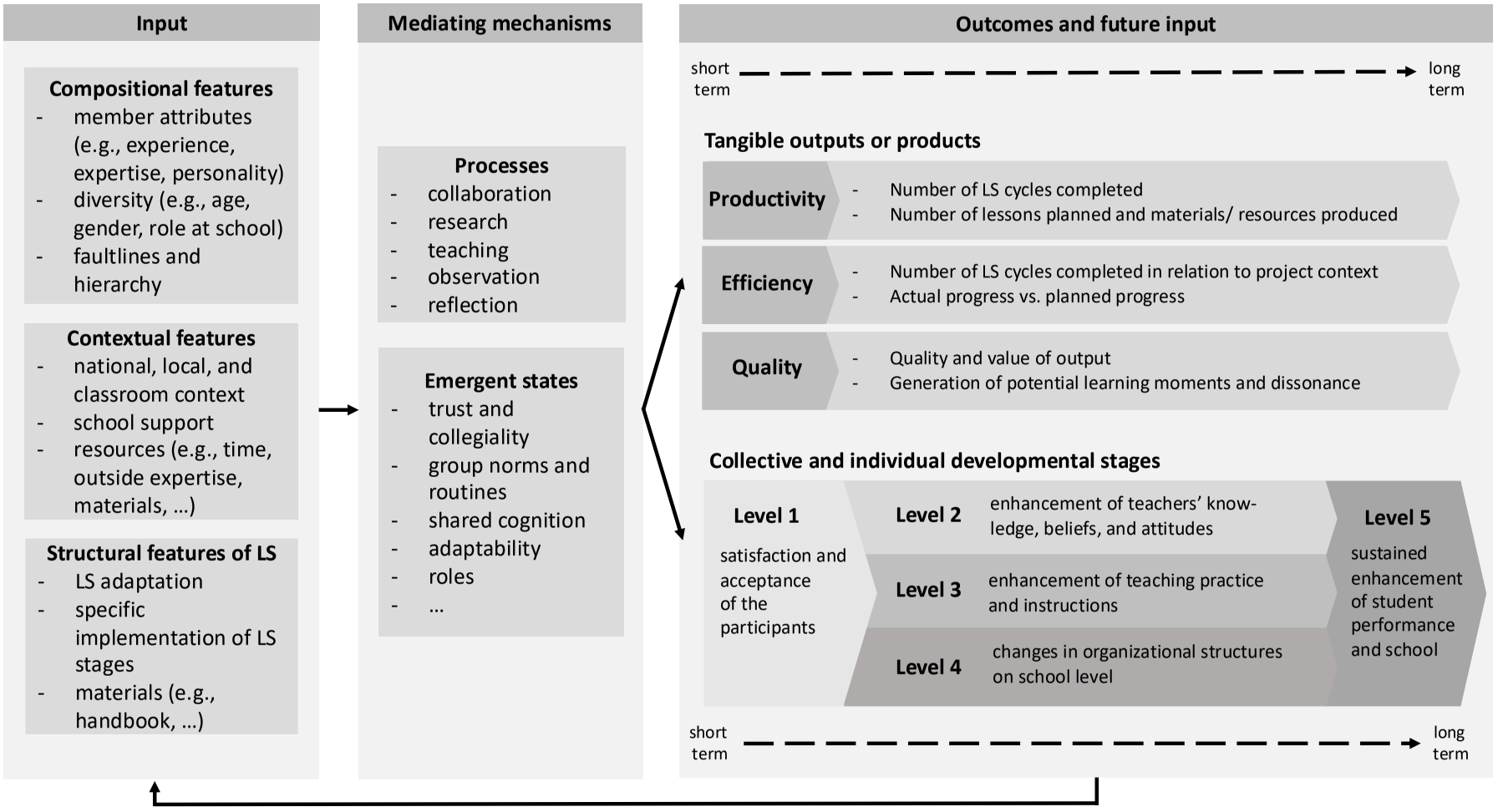
processes and mediating mechanisms. In the following, we describe each dimension (inputs, mediating mechanisms, outcomes and future inputs) and their component (Figure 2).

Input. We suggest the differentiation of three groups of input factors. Compositional features address the LS group itself and include member attributes (e.g., experience, personalities), diversity (demographic characteristics), and faultlines (i.e., factors that split the group into potential subgroups, Mathieu et al., 2019, p. 27). These aspects have been found to influence and predict group outcomes in studies on team effectiveness (Mathieu et al., 2019). Previous LS research also reports on several of these aspects as significant to a LS groups' work, i.e., members' teaching experience and LS experience and expertise (Bocala, 2015; Widjaja et al., 2017), their prior knowledge, attitudes, and beliefs (Bae et al., 2016; Norwich et al., 2016; Yoshida, 2012), their motivation and willingness to participate (Sjunnesson, 2020; Zhang, 2015), and hierarchical structures within groups (Chikamori et al., 2013; Lee & Madden, 2019). Compositional features therefore address who the individual teachers of a LS group are and how they find together as a team.

Turning to contextual features, the model includes the national, regional, and local school context of the LS group, the support they receive from their school, principal or project leaders, the status LS has within the school, as well as the classroom context and available resources (i.e., time, space, qualitative materials, access to external expertise and LS facilitators). The crucial role of these input factors have been repeatedly highlighted in the LS literature (e.g., Lee & Tan, 2020; Lim-Ratnam et al., 2019; Seleznyov et al., 2021; Xu & Pedder, 2014; Yoshida, 2012).

Figure 2

A conceptual model of continuous Professional Development through Lesson Study



As the third component of inputs, the model refers to the structural features of the PD, that is, the way in which LS was conducted. We identified this component (structural features of LS) as the most common starting point for existing LS models (Figure 1). This component can also be visualized in terms of the typical circular LS model (e.g., Lewis, 2009) that describes the core stages of study, plan, teach, and reflect. In our model, we suggest that three types of information are important in order to understand how the LS process are structured. The first concerns the specific type of LS that was adopted, such as Collaborative Lesson Research (e.g., Takahashi & McDougal, 2016), Research Lesson Study (e.g., Dudley et al., 2020), or Community-based Lesson Study (e.g., Yoshida et al., 2021). According to this type of LS, external experts or facilitators might assume diverging roles. The second type addresses the specific implementation of LS and its individual steps (i.e., How often did teachers meet? How was the planning stage conducted?). This feature is of interest to research, as the specific implementation of LS stages is likely to deviate even within a LS type. The third type of information relates to the LS materials that were used (e.g., handbook, observation protocols, etc.).

Mediating mechanisms. The next dimension concerns the means by which LS groups learn. Our model splits this dimension into processes and emergent states (Mathieu et al., 2019). Concerning processes, we have synthesized five processes that surface repeatedly across the LS literature and are also increasingly talked about in terms of "skills" teachers need in order to conduct LS, or alternatively, develop through engaging in LS: collaborating, researching, teaching, observing, reflecting, and. First, LS is a collaboration-based activity and requires teachers to establish a shared goal and vision, set up norms, and move forward as a group (e.g., Cammarata & Haley, 2018; Quaresma & Da Ponte, 2021). As teachers then conduct research on their own practice, they need to develop research questions, hypothesize about findings, design lesson plans, and analyze observational data (e.g., Fernandez, 2002; Wolthuis et al., 2020). During this inquiry process, teachers conduct a systematic observation of student learning to produce data, which requires teachers to notice salient classroom events (e.g., Amador & Carter, 2018; Karlsen & Helgevold, 2019). Teachers also need to engage in critical and collaborative reflection at all stages of the LS process, and especially during the reflection stage (e.g., Callahan, 2019; Calleja & Formosa, 2020; Kager et al., 2022; Mynott, 2019). This process is also frequently discussed in the LS literature in terms of "dialogue" and "talk" (e.g., Warwick et al., 2016).

These processes are accompanied by, and give rise to, emergent states, that is, to dynamic group properties (Mathieu et al., 2019). Some emergent states, while not referred to as such, have already been highlighted in the literature. Khokhova (2018), for instance, talks about LS groups developing trust and a sense of collegiality, while others have discussed group norms and routines (Lewis et al., 2019), or the development of groups' orientation towards collaboration (e.g., Quaresma & Da Ponte, 2021; Skott & Møller, 2017). Mathieu et al. (2019) also list a group's shared cognition, adaptability, efficacy, and the development of roles for this dimension. While these aspects have not yet been widely documented in LS research, we hypothesize that they are also relevant to LS groups.

Outcomes and Future Inputs. In accordance with Mathieu et al. (2019), the model distinguishes between tangible outputs or products, and collective and individual developmental stages. Tangible outputs include three categories: the LS groups' productivity and efficiency, as well as the value or quality of their outputs and products. In this context, we use outputs to refer to new insights, ideas, or intentions that result from the LS work and exist in teachers' minds. Products, on the other hand, include concrete lesson plans, materials, and other resources developed or adapted by the LS group. A LS group's tangible outputs are not fixed in time, but they grow and progress each time that a team engages in a new LS cycle.

Turning to the developmental stages, we suggest to adopt Guskey's (2000) five levels as a heuristic to track how LS impact evolves. This is similar to Lipowsky and Rzejak (2015), however, we assume that levels 2, 3, and 4 (enhancement of teachers' knowledge, teaching practices, and changes in organizational structures) cannot be neatly separated into linear events. We instead hypothesize, as indicated in the model, that changes on these levels develop in a dynamic and parallel fashion. As LS groups finish a cycle, they develop their outcomes and then return to the starting point (input) for the next cycle. This starting point evolves with the group and will look slightly different for each cycle.

Through continuous cycles, LS groups grow their tangible outcomes and, ideally, proceed along the developmental stages. The model emphasizes that these stages might differ between individuals and as a group. Level 5 implies sustained changes in student performance. This is arguably the end-goal of every PD (Darling-Hammond et al., 2017), yet Seleznyov's (2019) review highlights that the majority of LS research has so far either neglected to examine this outcome level or investigated student outcomes after only one or two LS cycle. As Guskey and Yoon (2009) note, demonstrating a clear relationship between

any PD and student improvement is a challenge that requires rigorous research designs and thoughtful planning. While the current model does not suggest concrete ways to measure LS' impact on students, it puts us into a better position to gather evidence by conceptualizing LS as an iterative and long-term PD that yields outcomes over time.

How to use the proposed model of Lesson Study

In a last step, we delineate several ways in which the model can be used in research and practice. Specifically, we suggest that the model serves as both a roadmap and a tool for analysis and evaluation (Table 1).

Table 1

Recommendations for How the Model and it's Adapted Version Can Be Used

	Examples for researchers	Examples for practitioners
Roadmap	to describe local LS adaptions and tie them to a shared framework	to introduce schools and teachers to LS and communicate its goals and scope
	to connect LS to the wider field of PD and discourse on professional development	to support early goal setting and realistic expectations
	to derive conceptual coherence and a shared terminology for the field	to keep track of LS work, e.g., by revisiting the model after each cycle
Tool	to situate existing research on a shared conceptual grid and derive new areas of research interest	to pinpoint areas in which the group wants or needs to make improvements
	to help explain findings of empirical research studies	to find answers to why certain outcomes have not yet been reached
	to identify and assess LS outcomes and stages of development	to communicate successes and needs to school boards or project leaders

Concerning its' application in research, the model acts as a roadmap (Jaakkola, 2020) by describing LS's crucial components and assuming relationships between them. Specifically, we suggest that researchers can use the model as a shared point of reference when describing their specific LS work in research articles. The model does not presume a specific LS adaption, cultural context, or subject, but posits components (e.g., the component of structural features of LS, or the component of tangible outputs and products), which researchers can then

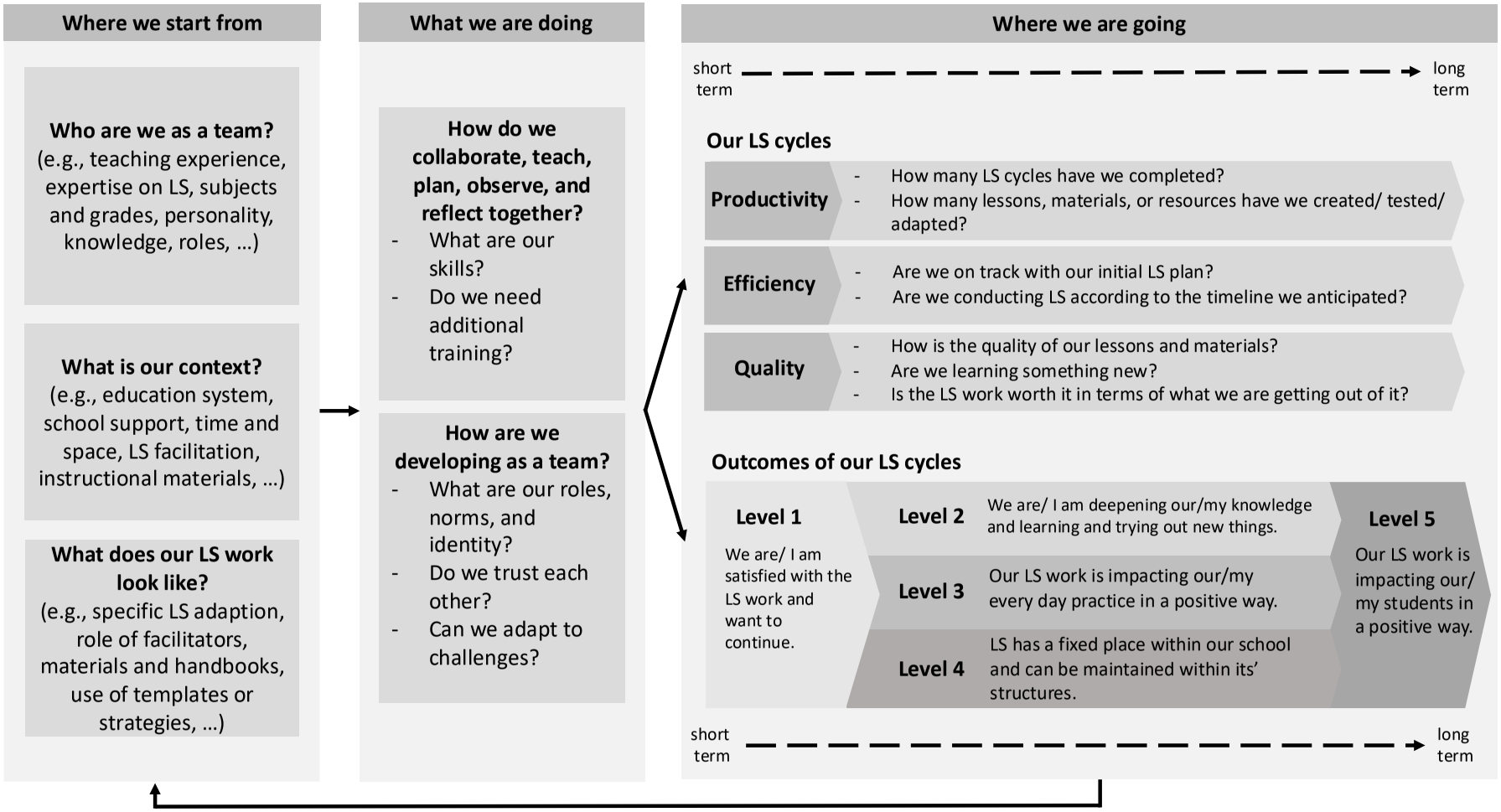
formulate and specify according to their LS work. By embedding descriptions of LS in the conceptual model, we could ensure the conceptual coherence between research studies and work towards the use of a shared terminology (i.e., by speaking about trust as a dynamic property of LS groups and situating it within the dimension of mediating mechanisms). Further, the model's description of LS outcomes could support the discourse in the field on using appropriate and, importantly, comparable outcome LS measures in research (e.g., Cheung & Wong, 2014; Seleznyov, 2019). Lastly, given that the model is abstract and does not presume specific PD elements, it could even be used to conceptualize Learning Study (e.g., Cheng & Lo, 2013) or similar continuous improvement methods and therefore enables us to better connect research on LS with the wider research field on continuous PD.

The model acts as a tool (Boylan et al., 2018), in so far that it can guide the analysis of how teachers learn through LS and aid the interpretation of research findings. Specifically, the model assumes relations between the three dimensions and their components factors. These relations can help us to better understand how, for example, a LS cycle develops or why certain LS groups seem more efficient or satisfied than others. While these relations have been examined in LS previous literature (e.g., Kitada, 2022), they have not yet been extensively tested. Empirical research on team effectiveness, however, indicates that input factors can explain and even predict outcomes; for instance, team members' expertise or their sense of how safe and confident they feel in a team can predict team performance (Mathieu et al., 2019). The model we propose can thus act as a tool to identify similar relevant relations that should be further analyzed and tested.

As the model is general, rather than derived from inductive analysis or through the application of a specific theory, it can be combined with various theoretical lenses or methodologies that are commonly used in LS research, such as sociocultural theory (e.g., Vygotsky, 1986) or cultural historical activity theory (e.g., Edwards, 2007). At this point it is also important to note that the proposed model is not meant to replace existing models, but to extend them and be used in combination with them. The model's input component of structural features of LS, for instance, refers to the surfaces features of LS, which are illustrated in the classic model of a LS process in the form of a cycle (e.g., Lewis et al., 2006). Our model does not aim to replace this cyclical LS model, but can be used as an elaboration of the cycle that takes multiple additional aspects of LS into account.

Figure 3

Adapted Version of the Lesson Study Model for Teachers' Own Assessment of Their Lesson Study Work



Turning to the model's application in practice, we share the view that LS is a teacher-led PD and as such, teachers need to be empowered in their role (e.g., Godfrey et al., 2019; Huang et al., 2016; Stigler & Hiebert, 2016). The adapted version of the model (Figure 3), which poses questions and avoids overly technical language, can achieve this in several ways. First, the model can support schools and teachers new to LS in understanding the scope of continuous improvement and developing realistic expectations concerning its outcomes. Clear goals from the onset further enable teachers to make visible their successes early on, which can provide encouragement to continue (Guskey, 2021). Likewise, the model can act as an initial how-to guide for project leaders or schools in regard to planning LS cycles and establishing short- and long-term goals, but also in finding together as a group and making space and time for LS work. This way, the model acts as a roadmap that accompanies LS groups from the beginning on.

We further suggest that the model can act as tool of analysis and evaluation guideline, allowing teachers to pinpoint or predict potential weaknesses in their own LS work. The idea is to provide teachers with a structure and language that supports them in voicing their own ideas about how LS can work for them, and which aspects need to be tweaked in order for LS to yield useful results. Darling-Hammond et al. (2017) emphasize the importance of credible means for teachers to evaluate their PD work. If practitioners know how to identify conditions needed to improve their outcomes, such as time, space, or additional expertise, this could support their communication of these needs to school boards, administrations, or policy makers (Darling-Hammond et al., 2017). For example, a group may find that they are productive and efficient in their LS work, but generally do not produce new insights or materials, thus doubting the value of LS. By systematically considering the various components of the model, the group could try to identify potential causes and solutions, i.e., they might require additional training in classroom observation, or additional study materials to develop qualitative lesson plans.

Conclusion, unresolved questions, and implications

The LS literature has grown and matured significantly in the past few decades. It has been pointed out that, for the field to move forward and engage in a coherent dialogue across the globe, a stronger theorization of LS and a shared conceptual framework are needed (e.g., Cheung & Wong, 2014; Stigler & Hiebert, 2016, Kager et al., 2022). In this paper, we have proposed a theory-informed model that can serve as such a conceptual framework. The model

advances existent LS models by adopting an IMOI-structure and outlining concrete inputs, mediating mechanisms, and two types of outcomes that develop over time. As Bryk (2015) notes, replacing PD programs with something new as soon as they seem ineffective is often just a short-term solution. Focusing instead on understanding how a PD works and why it does not yield the desired results is more likely to lead to continuous and long-lasting progress (Bryk, 2015). We consider the proposed model a crucial step towards viewing LS through the lens of improvement science, as it helps to increase our understanding of how to continuously improve various aspects of LS within a shared conceptual reference frame.

The proposed model nevertheless comes with limitations. Some issues remain unresolved, namely the question of how we can distinguish between inputs, mediating mechanisms and outcomes, how outcomes can be assessed in practice, and the model's empirical application.

The difficulty of placing certain factors within a specific dimension of the model is best illustrated by the example of (outside) expertise. Outside experts or LS facilitators could be reasonably placed within all dimensions and even within all three groups of input factors. For instance, expertise can be considered a compositional feature (i.e., a team members expertise and experience in facilitating LS), a contextual feature (i.e., the school's ability to organize external support), or a structural feature of LS (i.e., the specific role that an external expert assumes based on the type of LS and its concrete implementation). Facilitation could also be placed within the dimension of mediating mechanisms, if we consider it an additional process that supports teachers' learning. Lastly, expertise and knowledge of how to facilitate a LS team can also be viewed as a relevant outcome of LS. As Mynott and Michel (2022) note, however, research on facilitation in LS is currently still limited.

We therefore highlight that, despite bridging some gaps, the proposed model cannot yet provide answers to all questions. As a model, it remains an abstraction of the actual LS process. We argue, however, that the model offers a starting point for further research into the nuances of outside expertise and facilitation in LS, a topic that is currently still ambiguous in the research literature.

Another issue connected to this problem addresses the question of when a learning process becomes an outcome. Given that our model conceptualizes continuous PD through LS, we argue that outcomes are not final, but simply represent a certain stage in a dynamic continuous learning process. This means that a LS team can set their own goals for their LS work and evaluate their progress at selected stages in the process. The model aims to support

this continuous evaluation by conceptualizing outcomes of a LS process that serve as inputs for the subsequent process.

Second, the model does not specify ways in in which outcomes can be assessed, as this would be beyond the scope of this paper. We point, however, towards literature that has suggested various approaches to the assessment of PD and LS (Dudley et al., 2019; Godfrey et al., 2019; Guskey, 2000; Kennedy, 2005; Seleznyov, 2019), and suggest to explore ways in which they can inform LS research. We also argue that a stronger theorization of the individual processes included in the model can support the development of LS evaluation tools. For instance, different theoretical frameworks, such as dialogue and talk (e.g., Karlsen & Helgevold, 2019; Warwick et al., 2016), cognitive conflict (e.g., Calleja & Formosa, 2020; Mynott, 2019), or reflective stages (e.g., Kager et al., 2022) have been used to explain and analyze how teachers reflect critically in LS. Karlsen and Helgevold (2019) and Mynott (2019) effectively demonstrate that collaborative talk in LS does not guarantee that reflection will take place, but without the dialogic component reflection is unlikely to develop. As Kager et al. (2021) note, a deeper understanding of reflection is needed in the LS literature and a further exploration of these theoretical frameworks could be an avenue to better understand certain aspects of the model, such as processes and their quality, as well as the outcome category quality of tangible outputs.

Third, the model is, at this point, purely conceptual and has not yet been applied to research or real-life settings. We have put forward concrete recommendations for its usage and encourage to apply and test it rigorously across contexts. We view theorization as an iterative process (Davidoff et al., 2015; Weick, 1989) and thus conclude with the remark that the proposed model should be continually revised based on new empirical insights and optimized over time so that it can best serve the research community as a shared roadmap and tool of analysis and evaluation.

References

Adamson, B., & Walker, E. (2011). Messy collaboration: Learning from a Learning Study. *Teaching and Teacher Education*, *27*(1), 29–36. https://doi.org/10.1016/j.tate.2010.06.024

Akiba, M., Murata, A., Howard, C. C., & Wilkinson, B. (2019). Lesson study design features for supporting collaborative teacher learning. *Teaching and Teacher Education*, *77*, 352–365. https://doi.org/10.1016/j.tate.2018.10.012

Amador, J. M., & Carter, I. S. (2018). Audible conversational affordances and constraints of verbalizing professional noticing during prospective teacher Lesson Study. *Journal of Mathematics Teacher Education*, *21*(1), 5–34. https://doi.org/10.1007/s10857-016-9347-x

Arani, M. (2006). Transnational learning: The integration of Jugyou Kenkyuu into Iranian teacher training. In M. Matoba, K. Crawford, & M. R. S. Arani (Eds.), *Lesson Study: International Perspective on Policy and Practice* (pp. 37–75). Educational Science Publishing House.

Bae, L. C., Hayes, K. N., Seitz, J., O'Connor, D., & DiStefano, R. (2016). A coding tool for examining the substance of teacher professional learning and change with example cases from middle school science lesson study. *Teaching and Teacher Education*, *60*, 164–178. https://doi.org/10.1016/j.tate.2016.08.016

Bjuland, R., & Mosvold, R. (2015). Lesson study in teacher education: Learning from a challenging case. *Teaching and Teacher Education*, *52*, 83–90. https://doi.org/10.1016/j.tate.2015.09.005

Bocala, C. (2015). From experience to expertise: The development of teachers' learning in Lesson Study. *Journal of Teacher Education*, *66*(4), 349–362. https://doi.org/10.1177/0022487115592032

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher*, *33*(8), 3–15. https://doi.org/10.3102/0013189X033008003

Borko, H., Jacobs, J., & Koellner, K. (2010). Contemporary approaches to teacher professional development. In P. Peterson, E. Baker, & B. McGaw (Eds.), *International Encyclopedia of Education* (Volume 7, pp. 548-556). Elsevir.

Boylan, M., Coldwell, M., Maxwell, B., & Jordan, J. (2018). Rethinking models of professional learning as tools: A conceptual analysis to inform research and practice. *Professional Development in Education*, *44*(1), 120–139. https://doi.org/10.1080/19415257.2017.1306789

Bryk, A. S. (2015). 2014 AERA Distinguished lecture: Accelerating how we learn to improve. *Educational Researcher*, *44*(9), 467–477. https://doi.org/10.3102/0013189X15621543

Cajkler, W., Wood, P., Norton, J., & Pedder, D. (2013). Lesson Study: Towards a collaborative approach to learning in Initial Teacher Education? *Cambridge Journal of Education*, *43*(4), 537–554. https://doi.org/10.1080/0305764X.2013.834037

Callahan, C. (2019). Middle school geography teachers' professional development centered around historical photographs. *Journal of Social Studies Research*, *43*(4), 375–388. https://doi.org/10.1016/j.jssr.2018.11.003

Calleja, J., & Formosa, L. (2020). Teacher change through cognitive conflicts: The case of an art lesson study. *International Journal for Lesson & Learning Studies*, *9*(4), 383–395. https://doi.org/10.1108/IJLLS-05-2020-0028

Cammarata, L., & Haley, C. (2018). Integrated content, language, and literacy instruction in a Canadian French immersion context: A professional development journey. *International Journal of Bilingual Education and Bilingualism*, *21*(3), 332–348. https://doi.org/10.1080/13670050.2017.1386617

Canonigo, A. M. (2016). Using a non-coercive process to engage mathematics teachers in lesson study. *International Journal for Lesson and Learning Studies*, *5*(4), 329–347. https://doi.org/10.1108/IJLLS-02-2016-0004

Celik, A. O., & Guzel, E. B. (2020). How to improve a mathematics teacher's ways of triggering and considering divergent thoughts through Lesson Study. *International Electronic Journal of Mathematics Education*, *15*(3). https://doi.org/10.29333/iejme/8461

Cheng, E. C. K., & Lo, M. L. (2013). *Learning Study: Its origins, operationalisation, and implications* (OECD Education Working Papers No. 94). https://doi.org/10.1787/5k3wjp0s959p-en

Cheung, W., & 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*, *3*(2), 137–149. https://doi.org/10.1108/IJLLS-05-2013-0024

Chikamori, K., Ono, Y., & Rogan, J. (2013). A Lesson Study approach to improving a biology lesson. *African Journal of Research in Mathematics, Science and Technology Education*, *17*(1–2), 14–25. https://doi.org/10.1080/10288457.2013.826967

Chua, V. (2019). Going the distance: A Lesson study on deriving the distance formula. *International Journal for Lesson and Learning Studies*, *8*(2), 149–159.

https://doi.org/10.1108/IJLLS-08-2018-0052

Clarke, D., & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education*, *18*(8), 947–967. https://doi.org/10.1016/S0742-051X(02)00053-7

Coenders, F., & Verhoef, N. (2019). Lesson Study: Professional development (PD) for beginning and experienced teachers. *Professional Development in Education*, *45*(2), 217–230. https://doi.org/10.1080/19415257.2018.1430050

Da Ponte, J. P., Quaresma, M., & Mata-Pereira, J. (2022). Teachers' learning in lesson study: Insights provided by a modified version of the interconnected model of teacher professional growth. *ZDM – Mathematics Education*, 14. https://doi.org/10.1007/s11858-022-01367-1

Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective teacher professional development*. Learning Policy Institute. https://learningpolicyinstitute.org/product/teacher-prof-dev

Davidoff, F., Dixon-Woods, M., Leviton, L., & Michie, S. (2015). Demystifying theory and its use in improvement. *BMJ Quality & Safety*, *24*(3), 228–238. https://doi.org/10.1136/bmjqs-2014-003627

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, *38*(3), 181–199. https://doi.org/10.3102/0013189X08331140

Dick, L. K., Appelgate, M. H., Gupta, D., & Soto, M. M. (2022). Continuous improvement Lesson Study: A model of MTE professional development. *Mathematics Teacher Educator*, *10*(2), 19. https://doi.org/10.5951/MTE.2020.0077

Driskell, J. E., Salas, E., & Driskell, T. (2018). Foundations of teamwork and collaboration. *American Psychologist*, *73*(4), 334–348. https://doi.org/10.1037/amp0000241

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*, *34*, 107–121. https://doi.org/10.1016/j.tate.2013.04.006

Dudley, P., Xu, H., Vermunt, J. D., & Lang, J. (2019). Empirical evidence of the impact of lesson study on students' achievement, teachers' professional learning and on institutional and

system evolution. *European Journal of Education*, *54*(2), 202–217. https://doi.org/10.1111/ejed.12337

Dudley, P., Pratt, M., Gilbert, C., Abbey, J., Lang, J., & Bruckdorfer, H. (2020). Cross-school 'close-to-practice' action research, system leadership and local civic partnership re-engineering an inner-city learning community. *London Review of Education*, *18*(3), 390–407. https://doi.org/10.14324/LRE.18.3.05

Edwards, A. (2007). Relational agency in professional practice: A CHAT analysis. *Actio: An International Journal of Human Activity Theory*, *1*(3), 1–17. http://hdl.handle.net/10112/7572

Elliott, J. (2012). Developing a science of teaching through lesson study. *International Journal for Lesson and Learning Studies*, *1*(2), 108–125. https://doi.org/10.1108/20468251211224163

Elliott, J. (2019). Quality criteria for lesson and learning studies as forms of action research. *International Journal for Lesson and Learning Studies*, *9*(1), 11–17. https://doi.org/10.1108/IJLLS-02-2019-0018

Fernandez, C. (2002). Learning from Japanese Approaches to professional development: The case of Lesson Study. *Journal of Teacher Education*, *53*(5), 393–405. https://doi.org/10.1177/002248702237394

Festinger, L. (1962). Cognitive dissonance. *Scientific American*, *207*(4), 93–106. https://www.jstor.org/stable/10.2307/24936719

Fujii, T. (2014). Implementing Japanese Lesson Study in foreign countries: Misconceptions revealed. *Mathematics Teacher Education and Development*, *6*(1), 65–83. https://doi/10.3316/aeipt.205654

Godfrey, D., Seleznyov, S., Anders, J., Wollaston, N., & Barrera-Pedemonte, F. (2019). A developmental evaluation approach to lesson study: Exploring the impact of Lesson Study in London schools. *Professional Development in Education*, *45*(2), 325–340. https://doi.org/10.1080/19415257.2018.1474488

Groves, S., Doig, B., Vale, C., & Widjaja, W. (2016). Critical factors in the adaptation and implementation of Japanese Lesson Study in the Australian context. *ZDM: The International Journal on Mathematics Education*, *48*(4), 501–512. https://doi.org/10.1007/s11858-016-0786-8

Guskey, T. R. (2000). *Evaluating professional development*. Corwin Press.

Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching*, *8*(3), 381–391. https://doi.org/10.1080/135406002100000512

Guskey, T. R. (2009). Closing the knowledge gap on effective professional development. *Educational Horizons*, *87*(4), 224–233. https://www.jstor.org/stable/42923773

Guskey, T. R. (2021). Professional learning with staying power. *Educational Leadership*, *78*(5), 54–59. https://uknowledge.uky.edu/edp_facpub/52

Guskey, T. R., & Yoon, K. S. (2009). What works in professional development? *Phi Delta Kappan*, *7*, 495–500. https://doi.org/10.1177%2F003172170909000709

Gutierez, S. B. (2016). Building a classroom-based professional learning community through Lesson Study: Insights from elementary school science teachers. *Professional Development in Education*, *42*(5), 801–817. http://dx.doi.org/10.1080/19415257.2015.1119709

Hackman, J. R. (2012). From causes to conditions in group research: Causes and conditions. *Journal of Organizational Behavior*, *33*(3), 428–444. https://doi.org/10.1002/job.1774

Hackman, J. R., & Morris, C. G. (1975). Group tasks, group interaction process, and group performance effectiveness: A review and proposed integration. *Advances in Experimental Social Psychology*, *8*, 45–99. https://doi.org/10.1016/S0065-2601(08)60248-8

Hadfield, M., & Jopling, M. (2016). Problematizing lesson study and its impacts: Studying a highly contextualised approach to professional learning. *Teaching and Teacher Education*, *60*, 203–214. https://doi.org/10.1016/j.tate.2016.08.001

Hargreaves, E. (2005). Assessment for learning? Thinking outside the (black) box. *Cambridge Journal of Education*, *35*(2), 213–224. https://doi.org/10.1080/03057640500146880

Huang, R., Gong, Z., & Han, X. (2016). Implementing mathematics teaching that promotes students' understanding through theory-driven Lesson Study. *ZDM: The International Journal on Mathematics Education*, *48*(4), 425–439. https://doi.org/10.1007/s11858-015-0743-y

Ilgen, D. R., Hollenbeck, J. R., Johnson, M., & Jundt, D. (2005). Teams in organizations: From input-process-output models to IMOI models. *Annual Review of Psychology*, *56*(1), 517–543. https://doi.org/10.1146/annurev.psych.56.091103.070250

Isoda, M. (2015). The science of lesson study in the problem solving approach. In M. Inprasitha, M. Isoda, P. Wang-Inverson, & B. H. Yeap (Eds.), *Lesson study: Challenges in mathematics education* (pp. 81–108). World Scientific. https://doi.org/10.1142/9789812835420_0006

Jaakkola, E. (2020). Designing conceptual articles: Four approaches. *AMS Review*, *10*(1–2), 18–26. https://doi.org/10.1007/s13162-020-00161-0

Joubert, J., Callaghan, R., & Engelbrecht, J. (2020). Lesson Study in a blended approach to support isolated teachers in teaching with technology. *ZDM: The International Journal on Mathematics Education*, *52*(5), 907–925. https://doi.org/10.1007/s11858-020-01161-x

Kager, K., Jurczok, A., Bolli, S., & Vock, M. (2022). "We were thinking too much like adults": Examining the development of teachers' critical and collaborative reflection in lesson study discussions. *Teaching and Teacher Education*, *113*(103683), 1–13. https://doi.org/10.1016/j.tate.2022.103683

Kager, K., Kalinowski, E., Jurczok, A., & Vock, M. (2022). *A systematic review of transparency in Lesson Study research: How do we report on the observation and reflection stages?.* Manuscript submitted for publication.

Karlsen, A. & Helgevold, N. (2019). Lesson Study: Analytic stance and depth of noticing in post-lesson discussions. *International Journal for Lesson and Learning Studies*, *8*(4), 290–304. https://doi.org/10.1108/IJLLS-04-2019-0034

Kennedy, A. (2005). Models of continuing professional development: A framework for analysis. *Journal of In-Service Education*, *31*(2), 235–250. https://doi.org/10.1080/13674580500200277

Khokhotva, O. (2018). Lesson Study in Kazakhstan: Case study of benefits and barriers for teachers. *International Journal for Lesson and Learning Studies*, *7*(4), 250–262. https://doi.org/10.1108/IJLLS-04-2018-0021

King, F. (2014). Evaluating the impact of teacher professional development: An evidence-based framework. *Professional Development in Education*, *40*(1), 89–111. https://doi.org/10.1080/19415257.2013.823099

Kitada, Y. (2022). Teacher Agency in the Modification of Japanese Lesson Study in the United States. *Educational Studies in Japan*, *16*(0), 45–57. https://doi.org/10.7571/esjkyoiku.16.45

Korthagen, F. (2016). Inconvenient truths about teacher learning: Towards professional development 3.0. *Teachers and Teaching*, 1–19. https://doi.org/10.1080/13540602.2016.1211523

Larssen, D., Cajkler, W., Mosvold, R., Bjuland, R., Helgevold, N., Fauskanger, J., Wood, P., Baldry, F., Jakobsen, A., Bugge, H. E., Næsheim-Bjørkvik, G., & Norton, J. (2018). A literature review of lesson study in initial teacher education: Perspectives about learning and

observation. *International Journal for Lesson and Learning Studies*, *7*(1), 8–22. https://doi.org/10.1108/IJLLS-06-2017-0030

Lee, L., & Tan, S. (2020). Teacher learning in Lesson Study: Affordances, disturbances, contradictions, and implications. *Teaching and Teacher Education*, *89*, 102986. https://doi.org/10.1016/j.tate.2019.102986

Lee, V., & Madden, M. (2019). "We're in this together": Principals and teachers as partners and learners in Lesson Study. *NASSP Bulletin*, *103*(1), 51–64. https://doi.org/10.1177/0192636519826717

Lewis, C. (2009). What is the nature of knowledge development in Lesson Study? *Educational Action Research*, *17*(1), 95–110. https://doi.org/10.1080/09650790802667477

Lewis, C. (2015). What is improvement science? Do we need it in education? *Educational Researcher*, *44*(1), 54–61. https://doi.org/10.3102/0013189X15570388

Lewis, C. (2016). How does lesson study improve mathematics instruction? *ZDM*, *48*(4), 571–580. https://doi.org/10.1007/s11858-016-0792-x

Lewis, C., Perry, R., & Hurd, J. (2009). Improving mathematics instruction through lesson study: A theoretical model and North American case. *Journal of Mathematics Teacher Education*, *12*(4), 285–304. https://doi.org/10.1007/s10857-009-9102-7

Lewis, C., Friedkin, S., Emerson, K., Henn, L., & Goldsmith, L. (2019). How does Lesson Study work? Toward a theory of Lesson Study process and impact. In R. Huang, A. Takahashi, & J. P. da Ponte (Eds.), *Theory and practice of Lesson Study in mathematics* (pp. 13–37). Springer International Publishing. https://doi.org/10.1007/978-3-030-04031-4_2

Lewis, C., & Perry, R. (2014). Lesson Study with mathematical resources: A sustainable model for locally-led teacher professional learning. *Mathematics Teacher Education and Development*, *16*(1), 22-42. https://doi.org/10.3316/aeipt.205652

Lewis, C., & Perry, R. (2017). Lesson Study to scale up research-based knowledge: A randomized, controlled trial of fractions learning. *Journal for Research in Mathematics Education*, *48*(3), 261–299. https://doi.org/10.5951/jresematheduc.48.3.0261

Lewis, C., Perry, R., Hurd, J., & O'Connell, M. P. (2006). Lesson Study comes of age in North America. *Phi Delta Kappan*, *88*(4), 273–281. https://doi.org/10.1177/003172170608800406

Lewis, C., Perry, R., & Murata, A. (2006). How should research contribute to instructional improvement? The case of Lesson Study. *Educational Researcher*, *35*(3), 3–14.

https://doi.org/10.3102/0013189X035003003

Lewis, J. M., Fischman, D., Riggs, I., & Wasserman, K. (2013). Teacher learning in Lesson Study. *The Mathematics Enthusiast*, *10*(3), 538–619. https://scholarworks.umt.edu/tme/vol10/iss3/5

Lim, C. S., Kor, L. K., & Chia, H. M. (2016). Revitalising mathematics classroom teaching through Lesson Study (LS): A Malaysian case study. *ZDM: The International Journal on Mathematics Education*, *48*(4), 485–499. https://doi.org/10.1007/s11858-016-0779-7

Lim-Ratnam, C. T.-L., Lee, C. K.-E., Jiang, H., & Sudarshan, A. (2019). Lost in adaptation? Issues of adapting Japanese Lesson Study in non-Japanese contexts. *Educational Research for Policy and Practice*, *18*(3), 263–278. https://doi.org/10.1007/s10671-019-09247-4

Lipowsky, F. (2014). Theoretische Perspektiven und empirische Befunde zur Wirksamkeit von Lehrerfortund -weiterbildung [Theoretical perspectives and empirical findings on the effectiveness of teacher professional development]. In E. Terhart, H. Bennewitz, & M. Rothland (Eds.), *Handbuch der Forschung zum Lehrerberuf* (2nd edition, pp. 511–541). Waxmann.

Lipowsky, F., & Rzejak, D. (2015). Key features of effective professional development programmes for teachers. *Ricercazione*, *7*(2), 27-51.

Marks, M. A., Mathieu, J. E., & Zaccaro, S. (2001). A temporally based framework and taxonomy of team processes. *The Academy of Management Review*, *26*(3), 356–376. https://www.jstor.org/stable/259182

Mathieu, J. E., Gallagher, P. T., Domingo, M. A., & Klock, E. A. (2019). Embracing complexity: Reviewing the past decade of team effectiveness research. *Annual Review of Organizational Psychology and Organizational Behavior*, *6*(1), 17–46. https://doi.org/10.1146/annurev-orgpsych-012218-015106

Mathieu, J. E., & Gilson, L. L. (2012). Criteria issues and team effectiveness. In S. W. Kozlowski (Ed.), *The Oxford Handbook of Organizational Psychology* (Volume 2). OUP USA. https://doi.org/10.1093/oxfordhb/9780199928286.013.0027

Moss, J., Hawes, Z., Naqvi, S., & Caswell, B. (2015). Adapting Japanese Lesson Study to enhance the teaching and learning of geometry and spatial reasoning in early years classrooms: A case study. *ZDM: The International Journal on Mathematics Education*, *47*(3), 377–390. https://doi.org/10.1007/s11858-015-0679-2

Murata, A. (2011). Introduction: Conceptual overview of Lesson Study. In L. C. Hart, A. S.

Alston, & A. Murata (Eds.), *Lesson Study Research and Practice in Mathematics Education* (pp. 1–12). Springer Netherlands. https://doi.org/10.1007/978-90-481-9941-9_1

Mynott, J. P. (2019). Lesson study outcomes: A theoretical model. *International Journal for Lesson and Learning Studies*, *8*(2), 117–134. https://doi.org/10.1108/IJLLS-08-2018-0057

Mynott, J. P., & Michel, D. (2022). The invisible leader: Facilitation in Lesson Study. *Educational Process International Journal*, *11*(3), 48–61. https://dx.doi.org/10.22521/edupij.2022.113.3

Norwich, B., Koutsouris, G., Fujita, T., Ralph, T., Adlam, A., & Milton, F. (2016). Exploring knowledge bridging and translation in lesson study using an inter-professional team. *International Journal for Lesson and Learning Studies*, *5*(3), 180–195. https://doi.org/10.1108/IJLLS-02-2016-0006

Norwich, B., & 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*, *34*, 162–173. https://doi.org/10.1016/j.tate.2013.04.012

Norwich, B., & Ylonen, A. (2015). A design-based trial of Lesson Study for assessment purposes: Evaluating a new classroom based dynamic assessment approach. *European Journal of Special Needs Education*, *30*(2), 253–273. https://doi.org/10.1080/08856257.2015.1009702

Opfer, V. D., & Pedder, D. (2011). Conceptualizing teacher professional learning. *Review of Educational Research*, *81*(3), 376–407. https://doi.org/10.3102/0034654311413609

Quaresma, M., & Da Ponte, J. P. (2021). Developing collaborative relationships in lesson study. *PNA. Revista de Investigación En Didáctica de La Matemática*, *15*(2), 93–107. https://doi.org/10.30827/pna.v15i2.16487

Schipper, T., Goei, S. L., de Vries, S., & van Veen, K. (2017). Professional growth in adaptive teaching competence as a result of Lesson Study. *Teaching and Teacher Education*, *68*, 289–303. https://doi.org/10.1016/j.tate.2017.09.015

Schipper, T. M., de Vries, S., Goei, S. L., & van Veen, K. (2020). Promoting a professional school culture through lesson study? An examination of school culture, school conditions, and teacher self-efficacy. *Professional Development in Education*, *46*(1), 112–129. https://doi.org/10.1080/19415257.2019.1634627

Seleznyov, S. (2018). Lesson study: An exploration of its translation beyond Japan.

International Journal for Lesson and Learning Studies, *7*(3), 217–229. https://doi.org/10.1108/IJLLS-04-2018-0020

Seleznyov, S. (2019). Lesson study beyond Japan: Evaluating impact. *International Journal for Lesson and Learning Studies*, *8*(1), 2–18. https://doi.org/10.1108/IJLLS-09-2018-0061

Seleznyov, S., Goei, S. L., & Ehren, M. (2021). International policy borrowing and the case of Japanese Lesson Study: Culture and its impact on implementation and adaptation. *Professional Development in Education*, 1–15. https://doi.org/10.1080/19415257.2021.1973069

Sjunnesson, H. (2020). Initializing phase of lesson study: Communication a special didactic tool in mathematics. *International Journal for Lesson and Learning Studies*, *9*(3), 261–275. https://doi.org/10.1108/IJLLS-02-2020-0007

Skott, C. K., & Møller, H. (2017). The individual teacher in lesson study collaboration. *International Journal for Lesson and Learning Studies*, *6*(3), 216–232. https://doi.org/10.1108/IJLLS-10-2016-0041

Stake, R., & Schwandt, T. (2006). On discerning quality in evaluation. In I. Shaw, M. Mark, & J. Greene (Eds.), *Handbook of Evaluation* (pp. 405–418). SAGE Publications Ltd. https://dx.doi.org/10.4135/9781848608078.n18

Stigler, J. W., & Hiebert, J. (2016). Lesson Study, improvement, and the importing of cultural routines. *ZDM: The International Journal on Mathematics Education*, *48*(4), 581–587. https://doi.org/10.1007/s11858-016-0787-7

Takahashi, A., & McDougal, T. (2016). Collaborative lesson research: Maximizing the impact of lesson study. *ZDM*, *48*(4), 513–526. https://doi.org/10.1007/s11858-015-0752-x

Vanblaere, B., & Devos, G. (2021). Learning in collaboration: Exploring processes and outcomes. In Oude Groote Beverborg, T. Feldhoff, K. Maag Merki, & F. Radisch (Eds.), *Concept and design developments in school improvement research: Longitudinal, multilevel and mixed methods and their relevance for educational accountability* (pp. 197–218). Springer International Publishing.

Vrikki, M., Warwick, P., Vermunt, J. D., Mercer, N., & Van Halem, N. (2017). Teacher learning in the context of Lesson Study: A video-based analysis of teacher discussions. *Teaching and Teacher Education*, *61*, 211–224. https://doi.org/10.1016/j.tate.2016.10.014

Vygotsky, L. S. (1986). *Thought and language.* Massachusetts Institute of Technology.

Warwick, P., Vrikki, M., Vermunt, J. D., Mercer, N., & van Halem, N. (2016). Connecting

observations of student and teacher learning: An examination of dialogic processes in Lesson Study discussions in mathematics. *ZDM*, *48*(4), 555–569. https://doi.org/10.1007/s11858-015-0750-z

Watanabe, T., Takahashi, A., & Barham, A. I. (2019). Implementing school-wide collaborative lesson research in Qatar. *Journal of Institutional Research South East Asia*, *17*(2), 47–70. http://www.seaairweb.info/journal/articles/JIRSEA_v17_n02/JIRSEA_v17_n02_All.pdf#page=58

Weick, K. E. (1989). Theory construction as disciplined imagination. *Academy of Management Review*, *14*, 516–531. https://doi.org/10.5465/amr.1989.4308376

Widjaja, W., Vale, C., Groves, S., & Doig, B. (2017). Teachers' professional growth through engagement with Lesson Study. *Journal of Mathematics Teacher Education*, *20*(4), 357–383. https://doi.org/10.1007/s10857-015-9341-8

Wolthuis, F., Hubers, M. D., de Vries, S., & van Veen, K. (2020). More than mundane matters: An exploration of how schools organize professional learning teams. *International Journal of Leadership in Education*. https://doi.org/10.1080/13603124.2020.1790668

Xu, H., & Pedder, D. (2014). Lesson Study: An international review of the research. In *Lesson Study: Professional Learning for Our Time*. https://doi.org/10.4324/9780203795538

Ylonen, A., & Norwich, B. (2013). Professional learning of teachers through a lesson study process in England: Contexts, mechanisms and outcomes. *International Journal for Lesson and Learning Studies*, *2*(2), 137–154. https://doi.org/10.1108/20468251311323388

Yoshida, M. (2012). Mathematics Lesson Study in the United States: Current status and ideas for conducting high quality and effective lesson study. *International Journal for Lesson and Learning Studies*, *1*(2), 140–152. https://doi.org/10.1108/20468251211224181

Yoshida, N., Sugita, H., Kumai, S., & Fukuda, A. (2021). Lesson Study with multiple stakeholder: Community-based Lesson Study. In J. Kim, N. Yoshida, S. Iwata, & H. Kawaguchi (Eds.), *Lesson Study-Based Teacher Education: The Potential of the Japanese Appraoch in Global Settings* (pp. 183–198). Routledge.

Zhang, Y. (2015). Sustaining Lesson Study in schools with positive peer leadership: A case study in Hong Kong. *International Journal for Lesson and Learning Studies*, *4*(2), 140–154. https://doi.org/10.1108/IJLLS-07-2014-0018

Part III

Discussion

Part III: Discussion

This dissertation set out to analyze how the research field on LS has progressed over the past two decades and to pinpoint directions for future research. Systematic research on LS first appeared around the turn of the 21st century after LS was transferred from Japan to other contexts, most notably to the US. The emergence of the research field was driven, at least in part, by scholars' ambition to implement LS as a sustainable approach to teachers' PD in these new contexts. As Maddux and Cummings (2004) describe, educational interventions often fade away before they can be studied, improved, and adapted. Against this backdrop, Lewis et al. (2006) identified three research areas that needed to be addressed in order to prevent this faddism in the US context. These research areas—or critical research needs—were the development of a rich descriptive knowledgebase, the investigation of how teachers learn through LS, and the use of design-based research cycles to adapt, study, evaluate, and improve LS.

Almost two decades have passed since the formulation of these research needs. Taking stock of the current state of the field of LS, this dissertation found that LS can be considered a global success that has proven itself to be more than a "fad". The Japanese approach to teachers' PD has been adapted to educational systems around the world and is currently practiced in over forty countries (Yoshida et al., 2021), it is recognized as a promising PD approach within the wider field of education (Yurkofsky et al., 2020), and has brought together a large international community of scholars, educators, and policymakers. The cycle of faddism (Maddux & Cummings, 2004), however, is not always broken on a national and local level. Outside of Japan, LS frequently remains bound to projects or specific scholars (Huang & Shimizu, 2016) and empirical findings indicate that teachers are not always convinced that LS is beneficial to their practice or motivated to continue their LS work (Bjuland & Mosvold, 2015; Brown et al., 2016; Parks, 2008). This means that, while LS can be called an international success, the need to better understand how to implement effective LS sustainably in national and local settings remains of critical importance in the field.

The fact that this issue has not yet been sufficiently addressed or advanced is demonstrated by this dissertation's review of the LS literature. Specifically, the synthesis of reviews found that while a rich knowledge base on LS has been accumulated and a host of models to explain teacher learning through LS have been utilized in the literature, several key problems hinder further progress. These problems were 1) the frequent lack of comparable and replicable

descriptions of the LS intervention in publications, (2) the incoherent use or lack of use of theoretical frameworks to explain teacher learning through LS, (3) the inconsistent use of terminology and concepts, and (4) the lack of scientific rigor in research studies and of established ways or tools to measure the effectiveness of LS. In an effort to make progress on some of these issues, three studies were conducted as part of this dissertation.

The next chapter provides an overview of the key findings of these three studies. The subsequent chapter then discusses how these findings advance current research needs and derives implications for future research. The last two chapters acknowledge the limitations to this research and address how this dissertation contributes to the discourse on LS in Germany. The last chapter offers a final conclusion to this dissertation.

11. Key Findings of the Dissertation

This chapter briefly summarizes the research objectives of each study and reports their key findings and contribution.

11.1. Key Findings of Study 1

The goal of the first study was twofold: to gain a better understanding of the opportunities for teacher learning in the context of the post-lesson discussion inn a LS cycle from a theoretical perspective, and to examine how such a theoretical approach to critical and collaborative reflection may be translated into practice. Using a mixed-method design, the post-lesson discussions of four LS teams were examined in regard to their depth of reflection in terms of reflective stages, and the respective trajectories through their reflective practice.

Concerning its theoretical contribution, the article proposed a theoretically grounded definition for critical and collaborative reflection in the context of LS. It further demonstrated how a concrete theoretical model, namely Korthagen's (1985, 2010) ALACT model, can be used to rethink the post-lesson discussion. This rethinking of the post-lesson discussion in terms of Korthagen's reflection stages allowed to anticipate moments in which teachers might experience cognitive conflict and develop these potential moments of learning.

Concerning its methodological contribution, the article's theorization of the reflection stage informed the development of a coding tool to analyze of the depth of reflection. This coding tool can be reused or adapted by other scholars for similar purposes. In addition, the article employed an analytic approach, the micro-diachronic analysis of transcripts (Pitzl, 2020), which was developed within a different research field. This indicates the value of using analysis methods across disciplines.

The findings showed that the four LS teams differed significantly in the trajectories through their reflective practice. These results corroborate previous accounts that have identified reflection as a dynamic process (Korthagen, 2016) with seamless transitions between reflective stages (Rodgers, 2002). Specifically, the case studies indicated that groups underwent mini-cycles of reflection (Slavit & Nelson, 2010), meaning that proposed solutions or insights were re-tested and adjusted through a further exploration of the topic. We also saw that the presence of facilitators and the use of a theory-based reflection protocol may support teachers in their reflection process, yet they are no guarantee for a critical discussion between teachers. Instead, the findings highlighted that teachers' skills to reflect critically and

in a group are key for an effective post-lesson discussion. The article therefore identified critical and collaborative reflection as a LS core skill that needs to be clearly defined, practiced, and routinized. This takes the discourse by Mynott (2019) on LS core skills a step further und discusses the underlying processes of teacher learning in connection with skills, such as collaboration and reflection.

11.2. Key Findings of Study 2

In the light of the findings from Study 1, the goal of Study 2 was to explore how the stages of observation and reflection are conducted around the world, as well as theorized and reported in LS research. In a systematic review, we assessed whether publications on LS with in-service teachers provide sufficient details about the practical implementation of these stages and theorize the processes of observation and reflection. The focus on the stages of observation and reflection was chosen based on previous research findings. Specifically, Larssen et al. (2018) indicated that the stage of observation is ill-described in the LS literature on initial teacher education. Further, Study 1 showed that the concept of reflection, which is categorically underdescribed in the field of education (Brown et al., 2021; Walsh & Mann, 2015), also lacks a clear definition in the LS literature. The study addressed the following research questions: (1) How transparent are LS articles in reporting the steps of observation and reflection?; and (2) which – if any – theoretical frameworks are used to conceptualize the steps of observation and reflection in LS?

Concerning its theoretical contribution, the article applied a framework for transparency in social sciences—Moravcsik's (2020) three dimensions of transparency—to LS research. By doing so, a framework was established that details which information concerning the observation and reflection stages in LS are essential in publications. This framework informed the development of a coding tool that can be used to assess how transparent articles are in detailing how teachers observed and reflected in LS. The coding tool has been preregistered on OSF (Kager et al., 2021) and can be reused by scholars.

The findings of the article confirmed a lack of transparency in LS publications with in-service teachers and provided information into the extent of this issue. Concerning research question 1, we saw that the concepts of observation and reflection were inadequately described in the vast majority of the reviewed 129 articles. The most transparently reported category was the role of outside expertise, while other categories, such as how teachers documented their LS

process or whom teachers observed, were consistently neglected. The analysis identified some problems that contributed to this low transparency in LS publications: the presumption of a shared understanding of concepts, the omission or scattering of information, and the undertheorization of concepts. Concerning research question 2, we found a similar trend, namely that only a slim minority of articles clearly defined and/or theorized the concepts of observation and reflection. The analysis of the few articles that did provide a theorization identified a number of useful theories for the conceptualization of teachers' learning processes in LS, such as cognitive conflict (e.g., Festinger, 1962), modes of teacher talk (e.g., Mercer, 2000), and professional noticing (van Es, 2011).

Concerning its practical contribution, the article offers a check list that can inform researchers in their decisions on which information needs to be communicated in an empirical research paper to make their LS implementation transparent and replicable.

11.3. Key Findings of Study 3

The objective of Study 3 was to advance Lewis et al.'s (2006) research need 2, namely the development of a conceptual model of LS (see chapter 3). The overarching research question that framed the article was whether we can develop a model of LS that systematically addresses the dimensions of inputs, processes, and outputs, can account for LS's iterative character, and can be transferred between various contexts. To this end, the article analyzed existing LS models, identified current limitations, and consulted research on PD and from organizational psychology in order to solve these limitations and develop a conceptual model for the field of LS.

The contribution of this article is mainly theoretical. The proposed conceptual model of LS aims to serve as a shared schematic framework for the field and act as roadmap and a tool of analysis for both scholars and educators. While the model has not yet been put to the test in an empirical setting, the article demonstrates that it is possible to develop a theory-based model of LS that addresses all dimensions of the I-P-O framework and conceptualizes LS outcomes that develop over time. Importantly, the model pinpoints processes that teachers engage in when participating in LS, namely collaboration, research, observation, and reflection. This advances the discourse on LS core skills, which concluded Study 1, and widens the scope from collaborative reflection to other key processes that enable teacher learning in the context of LS.

Concerning its practical contribution, the article offers an adapted version of the conceptual model that is aimed towards practitioners and educators. This adapted model is intended to serve as a tool with which teachers can evaluate a LS cycle and gradually adapt their LS work until the desired outcomes are reached. This practical model needs to be tested and refined in the field, so that it can serve as a tool, or inform the development of tools, to evaluate and improve LS.

12. Looking Forward: How These Findings Advance Current Research Needs

The starting point of this dissertation were Lewis et al.'s (2006) critical research needs. A synthesis of all literature and systematic reviews of LS was conducted to assess the progress the field has made on these research needs. This synthesis revealed a number of limitations that hinder progress on the critical research needs in LS. These problems were (1) the frequent lack of comparable and replicable descriptions of the LS intervention in publications, (2) the incoherent use or lack of use of theoretical frameworks to explain teacher learning through LS, (3) the inconsistent use of terminology and concepts, and (4) the lack of scientific rigor in research studies and of established ways or tools to measure the effectiveness of LS. Together, the findings of the three studies conducted as part of this dissertation have generated new knowledge about the nature or extent of these problems and revealed some reasons behind them. The next subsections discuss this progress, as well as implications for future research, by first addressing insufficient descriptions of the LS intervention (problem 1), the incoherent use of theoretical frameworks and terminology (problems 2 and 3), and finally the scientific rigor in LS research (problem 4).

12.1. Insufficient Descriptions of the LS Intervention

Several previous studies have established that publications on LS tend to lack informative descriptions of their LS intervention (e.g., Baumfield et al., 2022; Cheung & Wong, 2014; Larssen et al., 2018). The reviews by Saito (2012) and Larssen et al. (2018) implied that the stages of observation and reflection remain particularly ambiguous in the research literature. To date, these findings have been mainly discussed on the sidelines and in the context of initial teacher education (e.g., Baumfield et al., 2022; Larssen et al., 2018). The existing discourse has highlighted the consequences of these problem, such as implementation challenges (e.g., Akiba et al., 2019; Seleznyov et al., 2021) and the difficulty of synthesizing the evidence on LS's efficacy (e.g., Cheung & Wong, 2014; Willems & van den Bossche, 2019). The reasons behind these problems, as well as possible solutions, have been largely overlooked so far.

By explicitly addressing the issue of insufficient information on the LS intervention, both Study 1 and Study 2 have generated new insights into the issue. Study 1 found that the LS literature lacks a shared definition for a key stage in the LS process, namely teachers'

collaborative and critical reflection. Study 2, the systematic review, corroborated and extended Larssen et al.'s (2018) findings by showing that publications on LS with in-service teachers tend to omit explicit information about the stages of observation and reflection. Specifically, the majority of examined LS publications omitted some or all information concerning how teachers conducted their classroom observations, as well as how teachers interpreted the collected observations or other type of data (e.g., students' work sheets). In addition, most publications failed to clarify framework conditions concerning how, when, or for how long LS stages were executed. Together, Study 1 and 2 thus demonstrate which details about the observation and reflection stages are currently underreported in the literature.

These findings have several implications for future research. In order to enhance the quality of the knowledge base on LS and advance Research Need 1 (i.e., Development of a coherent knowledge base on LS), the research community needs to develop an awareness for the importance of research transparency. To aid this process, Study 2 proposes a checklist concerning which information relating to the observation and reflection stages needs to be made explicit in publications. The design of the checklist is informed by similar checklists that aim to increase transparency in research descriptions (Aguinis et al., 2018; Moher et al., 2009). The list recommends, for example, to describe how researchers and the LS team understood the observation and reflection process from a theoretical perspective, how teachers actually carried out their post-lesson discussion, and whether the LS team used specific materials or protocols during either stage that could be of interest to readers. The main objective of the checklist is to promote descriptions of the LS intervention that are informative, comprehensive, and that clarify how researchers or teachers made choices in the LS process that shaped the outcome of this process.

By investing some work into creating transparent and comparable descriptions of the LS intervention in publications, the research field can greatly increase the quality of the knowledge base on LS. Importantly, transparent descriptions enhance the ways in which both researchers and educators can engage with and build on research. To this end, Study 2 recommends several practices to make research findings and LS materials more accessible to the research field, such as providing templates of LS protocols in the appendix, or to store supplementary materials on journal websites or online repositories. Looking forward, we can expect the empirical knowledge base on LS to continue to mature across the globe. A coherent

and transparent knowledge base on LS will make it more likely that future systematic reviews of this knowledge base can meet high standards in their synthesis of research findings.

12.2. Incoherent Use of Theoretical Frameworks and Terminology

The problem of insufficient descriptions of the LS intervention is closely related to the problem of vague terminology. The co-occurrence map, presented in Study 2, illustrated the variety of terminology used in publications to refer to the reflection stage. This terminology remains, however, frequently undefined in the literature. Study 2 identified several possible explanations for this ambiguous language. Specifically, this lack of awareness for the need of explicit language might stem from the presumption that concepts such as *LS*, *teacher learning*, *observation,* and *reflection* have a universal understanding. As Aguinis et al. (2018) argue, this presumption of a mutual understanding is a known problem in social sciences. In the field of education, the concept of reflection has been notoriously undertheorized und underdescribed (Brown et al., 2021; Mann & Walsh, 2013). The findings of Study 1 and Study 2 emphasize that such a mutual understanding of observation and reflection does not currently exist in the LS field. The fact that a recent study by Mynott and O'Reilly (2022) found a comparable ambiguity in LS publications about teachers' collaboration suggests that there are likely additional terms and concepts in the LS literature that need to be defined more precisely and against a common backdrop.

The use of ambiguous language in LS publications appears to be the symptom of a missing theoretical foundation that could support the development of a coherent understanding of LS. In fact, concerns about the lack of theorization in LS have been raised by several scholars in the past (e.g., Larssen et al., 2018; Saito, 2012; Stigler & Hiebert, 2016). Study 2 confirmed that the majority of research articles included in the systematic review did not connect teachers' observations or reflections to any theoretical account. The few articles that did theorize these LS stages drew on various theoretical underpinnings. Concerning the observation stage, the review identified the concept of noticing (e.g., van Es & Sherin, 2002; see Karlsen & Helgevold, 2019) as a promising starting point for subsequent efforts to better understand teachers' learning opportunities during the research lesson. Concerning the reflection stage, the review determined that current theorization efforts subscribe to two different schools of thought. Some article drew on sociocultural learning theory (Bandura, 1969; Vygotsky, 1986), while others drew on cognitive aspects of learning (Festinger, 1962;

Harmon-Jones & Mills, 2019; Posner et al., 1982). As a result, two distinct conceptualizations of how teachers learn in LS have been established in the literature: through collaborative talk and dialogue, or through reflection and cognitive conflict (see Study 2). This finding demonstrates that multiple understandings of teacher learning exist in the LS research literature and further lends support to the recommendation given in Study 2 that theoretical foundations cannot be presumed, but need to be made clear in publications.

These findings imply that the knowledge base on LS could achieve more coherence, if publications used intelligible terminology and a common conceptualization of the LS process. To make progress on this issue, Study 3 proposes a conceptual framework of LS that aims to act as a reference point within the field. A conceptual model, according to Jaakkola (2020), tries to characterize a process or phenomenon by describing its components and how they interact with each other. This definition seems to be largely in line with Lewis et al.'s (2006) understanding that a LS model would "specify the relationships between lesson study's observable aspects and [teachers'] instructional progress" (p. 5). The proposed conceptual model of LS seeks to unify LS research by not presuming any cultural context, LS adaptation, or theory of learning. Instead, it provides a framework and specific terminology that research studies can use to make their conceptualization of LS explicit.

A strength of the proposed model is that it supports the evaluation of LS. As noted in the introduction of this dissertation, the evaluation of PD is connected to several challenges. Guskey (2017) argues that some of these challenges can be addressed by considering the desired goals before the implementation of PD starts. In the LS context, this would mean that stakeholders (e.g., school leaders, project leaders, teachers, financial backers) agree on what a successful implementation of LS would look like before the actual LS work begins. There are, however, no commonly agreed on indicators of success in the LS literature. Da Ponte (2017) remarks, for instance, that the goals of LS in the initial teacher education literature differ substantially between research studies. In general, the majority of research studies connect successful LS to an increase in teachers' professional learning, a change in teachers' everyday practice, and teachers' satisfaction with LS (Seleznyov, 2019). The ways in which empirical studies evaluate, measure, and interpret these outcomes, however, vary widely (Cheung & Wong, 2014; Willems & van den Bossche, 2019).

The conceptual model therefore provides concrete outcomes of LS over time that may support both research and practice. Concerning research, the model can facilitate the

development of comparable outcome measures and instigate a productive dialogue in the field concerning the best tools and techniques to collect the necessary data. Concerning practice, the model and the dimension of outcomes, in particular, can support stakeholders in considering their central goals early on and develop realistic expectations. Guskey (2017) also remarks that opinions on what makes PD successful and worth the effort may differ considerably between stakeholders. School leaders, administrators, or financial backers may expect data-based results (e.g., quantitative results, standardized tests), whereas teachers tend to consider their own evaluations of student learning more trustworthy (Guskey, 2017). To meet both needs, the conceptual model differentiates between tangible outcomes (e.g., Productivity, Efficiency, Quality) and outcome stages that accumulate over time (from teachers' satisfaction to a sustained enhancement of students' performance).

Looking forward, it is hoped that the conceptual model of LS may assist the field both prospectively and retrospectively (Radovic et al., 2018). Prospectively, the model can support researchers in identifying areas and components that require further investigation and ensure that researchers pursue similar research goals and employ comparable methods, tools, and instruments. Retrospectively, the model may be used as an "analytic lens" (Radovic et al., 2018, p. 33) to interpret, categorize, or group existing research. It has to be noted, however, that the model has not yet been tested empirically or in real-world settings and is still only conceptual at this time.

12.3. Scientific Rigor in LS Research

The synthesis of LS reviews indicated that only a small minority of studies on the effectiveness of LS are based on controlled designs (Seleznyov, 2019; Willems & van den Bossche, 2019). In addition, Baumfield et al. (2022) report that the majority of studies included in their review failed to make their research design and methods explicit. Research methodologies were not the primary focus of the research papers conducted as part of this dissertation, but the findings of Study 2 nevertheless support previous assessments and emphasize that there is room for improvement regarding this issue.

First, the large number of qualitative research studies on LS identified in Study 2 are in line with Seleznyov's (2019) claim that qualitative research plays a significant role in forming our knowledge base on LS. Second, Study 2 corroborates Baumfield et al.'s (2022) finding that many articles on LS do not specify their methodological approach. Study 2, which examined

publications on LS with in-service teachers, found that 37% of studies did not describe their methodological approach in terms other than *qualitative*. Third, only a small minority of articles included in Study 2 categorized their approach as either AR (*n* = 3) or DBR (*n* = 4). This finding contradicts earlier studies that indicated that the field of LS makes extensive use of AR (Baumfield et al., 2022; Yoshida et al., 2021). This discrepancy may be explained by the fact that publications in Study 2 were only counted as AR or DBR if they classified their own approach in that manner. This means that additional publications based on AR may have been included in Study 2, but they were not identified as such as they did not make this approach explicit or used divergent terminology.

This dissertation argues that this observed lack of rigor in research studies as well as in the descriptions of research methodologies is a symptom of the field's dispersed approach to investigating LS. As summarized in the introduction, the investigation of PD approaches is connected to several challenges, such as linking PD to a change in teachers' instructions and students' achievements (Guskey & Yoon, 2009), coming up with the necessary resources to conduct longitudinal and controlled studies (Bryk, 2015), and taking the complex and dynamic environment of schools into account (Yurkofsky et al., 2020). This complexity of educational settings has been identified as one of the reasons why the WWC paradigm (i.e., *What Works Clearinghouse*; Bryk, 2015) is gradually being replaced by the paradigm of improvement science (Yurkofsky et al., 2020). The approach of improvement science is consistent with Lewis et al.'s (2006) suggestion of the so-called local proof route to spread educational innovations. The local proof route aims to develop innovations locally and in collaboration with practitioners and to improve and continuously adapt the innovation through local data and feedback. This route of spreading, improving, and evaluating LS in accordance with research designs such as DBR and AR seems like a promising idea for the field of LS. These efforts, however, need to be adequately described in publications in order to contribute to a reliable and replicable knowledge base (Tinoca et al., 2022). As the findings of Study 2 confirm, this is currently not the case in the LS literature. In addition, the small number of articles that cited DBR or AR as their research methodology raises the suspicion that the field of LS has yet to realize the full potential of these methodologies.

These findings have several implications for research into LS. To make progress on these issues, the field should first of all strive towards conceptual clarity concerning DBR, AR, and related terms, and develop standards for reporting research methodologies in a transparent

manner. As an important step into this direction, Study 2 recommends that publications make a clear distinction between their research methodology on the one hand and their LS intervention on the other hand (e.g., by reporting them in separate sections and using unambiguous terminology).

Second, the findings of this dissertation imply that in order to respond to calls for more robust evidence for the efficacy of LS (e.g., Cheung & Wong, 2014; Rzejak, 2019; Seleznyov, 2019; Willems & van den Bossche, 2019), the issue of methodology and evaluation needs to be more openly discussed in the field. To make a start, this dissertation puts up three ideas for discussion. First, while the WWC paradigm and improvement science may represent two distinct approaches to the evaluation of PD, they do not necessarily exclude one other. As Maddux and Cummings (2004) note, the effectiveness of PD is frequently evaluated prematurely, which eliminates the opportunity to modify its steps and make the PD useful for a specific context. This is why, according to Lewis et al. (2006), PD concepts should be refined in their local context—for instance through design-based cycles—before they are put to the test. In other words, improvement science can aid in the successful transfer of a PD concept to a new setting. The outcome of this process might then be evaluated using more rigorous research designs (e.g., randomized-controlled trial). It is acknowledged that such a strategy that bridges the ideas of both paradigms would require substantial resources, time, and funding. Yet, given how challenging it is to transfer PD concepts successfully from research to practice, it may be a strategy worth considering.

Another strategy could be to deconstruct LS and test the impact of its individual stages and activities. Century and Cassata (2016) describe this strategy as an empirical approach to generating knowledge about the effectiveness of an innovation's individual features in multiple settings. They argue that if we view innovations as a combination of specific features, we may learn not only about the effectiveness of the whole innovation, but we generate valuable knowledge about the role of individual components that make up the innovation (Century & Cassata, 2016). These insights might then inform the investigation of other innovations that include these components.

This approach from implementation research seems particularly potent for research on LS. The empirical literature reports multiple ways in which the LS stages of observation and reflection—which can be viewed as specific components of an innovation—are executed. These studies also describe several factors that may diminish the success of these stages (e.g.,

Bjuland & Mosvold, 2015; Callahan, 2019; Karlsen & Helgevold, 2019), or specific interventions that may facilitate the process, such as specific protocols (e.g., Færøyvik Karlsen, 2019) or external expertise (e.g., Bae et al., 2016). By adopting the strategy described by Century and Cassata (2016), we could refine and test the components of obervation and reflection individually. As a first step, we could identify promising procedures for these components that have been reported in the LS literature and consult the wider literature on techniques for classroom observation or collaborative reflection. Study 1, for instance, employed two distinct approaches that could be tetsed and improved; namely the learning activity curve (Knoblauch, 2019) for the observation stage, and a reflection protocol based on Korthagen's (1985, 2010) ALACT model. These appraoches could serve as the basis for more clearly defined procedures that could, in a next step, be tested and improved individually across multiple schools or in initial teacher education.

Refining individual components of LS, instead of the whole intervention, poses several advantages. It would likely require less financial ressources and time commitment of schools than conducting a randomized-controlled trial. Including a larger sample, multiple conditions, and ensuring a controlled setting would thus become more feasible. In addition, the findings might be relevant to the wider field of education, as they would relate to the efficacy of a specific component. The components of classroom observations and collaborative reflection are pertinent to a range of research inquiries in the field.

The results of such studies would need to be carefully reintergrated into the context of LS. This would require a critically evaluation of the variables in the LS context that might interact and influence the protocols and procedures tested in isolation. According to conceptual model of LS, the variables that could challenge such a transfer include, for example, the compositional characteristics of the LS team (such as participants' reflection experiences) or states that develop over the LS process (e.g., trust and collegiality, group norms). Nevertheless, this strategy from implementation research could be an avenue for future efforts to maximize and verify the impact of LS.

The third strategy refers back to Lewis et al.'s (2016) suggestion to use design-based research cycles to investigate and hone LS across different contexts. The review of DBR and AR in the introduction of the dissertation highligthed the value of design-based cycles as a methodology to progressivly hone LS. However, the findings of this dissertation, and of Study 2 in particular, suggest that the field of LS has not yet exploited the full potential of DBR. One

goal of DBR is to identify those principals that contribute to the success of an innovation (Anderson & Shattuck, 2012). DBR studies that provide explict information on the context of their LS implementation and the design principals that they identified would be a substantial contribution to the field.

12.4. Summary of Implications for Future Research

The findings of the three studies presented in this dissertation make it possible to advance the research needs (see chapter 3) even further. This section provides a short synopsis of how current research needs can be addressed in future research.

Research Need 1. The first research need should remain the development of an international, coherent, and rigorous knowledge base on LS. The findings of this dissertation indicate several practices that can aid this process, as well as some areas of interest, that should be addressed more openly in the field. These practices include explicit descriptions of both the research methodology and the LS intervention in publications (e.g., through the use of the checklist proposed in Study 2). Another recommended practice is the clarification of terminology and concepts, such as observation and reflection, as well as of any theoretical underpinnings. In order to enhance the accessibility of research findings, this dissertation recommends the use of Open Science practices whenever possible, such as providing online supplementary materials. To make materials used in the context of LS more transparent and also reusable, it is suggested, for instance, to include examples or templates in a publication or its appendix, to cite or link LS sources and materials, and to refer readers to websites or repositories where resources, such as observation protocols, can be accessed. In addition, making data publicly available on online repositories would aid future reviews and meta-analyses.

Turning to areas of interest, this dissertation recommends to renew the debate by, for instance, Elliott (2019) and Hanfstingl et al. (2019), concerning the scope of LS and its parallels and differences with AR. A better understanding of how AR and DBR are currently used in LS research would also aid the clarification of terminology. Another area of interest that connects to AR is the collaboration between researchers and teachers in LS. This collaboration, which according to Study 2 can take on various forms, should be further explored in research, especially regarding the dual role of teachers as researchers. Another key area of interest remains the evaluation of LS and methodological considerations connected to the assessment

of continuous PD. Several strategies forward concerning how the field of LS may address these methodological challenges have been derived and are summarized in Table 4.

Research Need 2. The second research need should be updated from the development of a shared conceptual model of LS to the testing and improvement of this model. This dissertation suggests three ways forward. First, it is suggested to use the model retrospectively to assess the current knowledge base on LS and group existing research. For instance, the model proposes a separation between processes and mediating mechanisms. A review of the literature could help identify to what degree these two components have already been examined in the literature, as well as whether additional mediating mechanisms have already been explored and might need to be added to the model. Second, it is suggested to use the model prospectively to inform the design of future studies. Research could examine, for instance, the relationship between individual components of the model, such as the ways in which the compositional features of a LS team affect the emergence of mediating mechanisms. Similarly, research could assess short- and long-term LS outcomes by employing methods suggested by Guskey (2000) and traditionally used for the assessment of PD (i.e., pre- and post-tests, questionnaires, interviews, students' achievements, …). It would also be advisable to explore how other research fields examine and measure similar processes (e.g., team effectiveness research, organizational psychology) and adapt suitable ideas to the study of LS. Third, the conceptual model and its adapted version should be continuously improved based on new research insights and researchers' and educators' experiences with the model. As Davidoff et al. (2015) remark, developing a model is an iterative process. The versions of the model proposed in this dissertation are thus not intended as finished products but rather as prototypes.

Research Need 3. The third research need should remain the development of a shared understanding of DBR and the PDSA cycle. Study 2, examining self-reported research designs, was unable to support previous claims that AR, which relies on PDSA cycles, is widely used in the field of LS (e.g., Yoshida et al., 2021). However, it is possible that studies are employing research designs consistent with either AR or DBR, but they use different classifications or terminology. Looking forward, the field would benefit from a shared conceptualization of DBR and AR, a synthesis of how design-based cycles are currently used in LS research, and from paying closer attention to the tools and protocols that can guide DBR and AR. An additional

stepping stone forward would be to have an open debate on whether LS itself is a form of AR and on the implications of classifying LS as classroom research rather than as continuous PD.

Table 4 brings together the three perspectives explored in this dissertation—then, now, and looking forward. The first row reviews the research needs identified by Lewis et al. (2006). The second row summarizes the current research needs, which were identified in the context of this dissertation. The third row outlines the strategies and practices discussed in this chapter that can support the field in pursuing these research needs in the future.

Table 4

Overview of Critical Research Needs: Then, Now, and Looking Forward

	Research Need 1	Research Need 2	Research Need 3
Then Research needs in LS according to Lewis et al. (2006)	**Expansion of the Descriptive Knowledge Base of Japanese and U.S. Lesson Study** to expand the descriptive knowledge base on LS in an effort to describe LS's characteristics and determine adaptations pertinent to LS's implementation in US settings	**Explication of the Innovation Mechanism** to investigate the mechanisms through which teachers learn in LS and develop a model that represents these mechanisms as well as LS surface features	**Design-Based Research Cycles** to use design-based research cycles to improve LS adaptations and support theory-building
Now Research needs in LS derived from the current literature	**Development of an international, coherent, and rigorous knowledge base on Lesson Study** to build an international research base that - uses theory-based and rigorous approaches to study teacher learning - provides transparent descriptions of their research methodologies and LS intervention	**Development of a conceptual model of Lesson Study** to develop a theory-based conceptual model of LS that - can be used across contexts - systematically describes inputs, processes, and outputs - can be applied widely	**Development of a understanding of DBR** to capitalize on improvement methodologies by - establishing clear terminology - synthesizing current use and findings of DBR in LS - developing and testing protocols that can guide DBR
Looking forward Strategies to advance current research needs	Practices to generate, systematize, and engage with knowledge base: - informative and separate descriptions of research methodologies and the LS intervention - transparent terminology - and theoretical underpinnings - Open Science practices - rigorous methodologies to examine efficacy Areas to examine: - lines between DBR, AR, and LS - collaboration between researchers and educators - components of the conceptual model	Approaches to test and use the conceptual model: - use model to group existing research - use model to plan future research - examine relationship of individual components - examine short- and long-term LS outcomes Approaches to refine improve model: - adapt model based on research insights and feedback from teachers - adapt model to serve as evaluation tool to teachers	Strategies to develop a shared understanding of DBR: - synthesize how AR and DBR are currently used in the field - conduct well- described DBR and synthesize and compare design-principals - discuss LS as form of AR/ PD and the implications of these classification

13. Limitations to this Dissertation

Although this dissertation reports valuable progress on the research needs in LS, the findings are subject to a number of limitations and should be interpreted with caution. The limitations that affect each of the three research papers have already been reported in each respective study. This section nevertheless wants to highlight three of these limitations, as they also have a bearing on the dissertation's framework.

First, due to the author's language abilities, the assessment of the current research needs in the introduction of this dissertation rely on two scoping reviews that only considered research published in English or German. The systematic review (Study 2) employed even stricter inclusion criteria and only considered articles published in English. This inclusion criterion is recognized as a delimitation, as additional eligible studies and relevant international contribution might have been overlooked. In addition, this selection perhaps lent more weight to research conducted in English-speaking or Western nations.

Second, this dissertation focused heavily on the LS stages of observation and reflection. This focus was chosen based on previous studies that have identified these two stages as both crucial and challenging to the success of LS, as well as undertheorized and underdescribed in publications (e.g., Larssen et al., 2018; Saito, 2012). This focus does not imply, however, that the first two stages of the LS process are less important. In fact, the conceptual model (Study 3) suggests that all four LS stages rely on the processes of collaboration, research, observation, and reflection. Looking forward, this dissertation suggests to examine each of these processes, as well as how they translate into skills that teachers need to possess or develop during LS in order to maximize their learning.

Third, this dissertation proposes a conceptual model of LS that, at this point, is purely theoretical and has not yet been tested in the field. Study 3 suggests several ways for how the model can be applied by both researchers and educators. As a next step, the model should therefore be tested in various settings and iteratively improved. Interesting questions for this research would be to test whether—and if yes, to what extend—individual components of the model influence each other. This limitation also affects the adapted version of the conceptual model, which is intended to function as a tool for teachers to evaluate their own LS work. It is acknowledged that the current version of this adapted model is not the end-product, but rather a prototype that needs to be refined through teachers' feedback and their experience with the model. Lastly, the model includes concrete outcomes of LS, but does not propose

how these outcomes should be evaluated or measured. The model should be viewed as a shared foundation for a discussion on what tools or approaches can best serve the field in this effort.

14. Practical and Theoretical Implications for Lesson Study in Germany

This dissertation, written in the context of the LemaS-project (BMBF & KMK, 2016), aimed to generate knowledge about how teachers at German schools reflect critically together and, in addition, derive considerations for research on LS in Germany by taking stock of the international research field. This chapter briefly outlines the progress made on these two objectives.

Concerning the first objective, Study 1 provided novel insights into how teachers at German schools engage in critical and collaborative reflection in the context of LS. The four primary schools included in the analysis were purposefully chosen based on their contrastive features (Patton, 1990). The goal of this sampling was to gain a comprehensive picture of LS across different German school settings and to identify affordances as well as hindrances to teachers' reflection processes. The findings show that LS was successful in providing a platform for teachers to intensify their collaboration and reflect on how students learned during the research lesson. They also demonstrate, however, that stakeholders who aim to implement LS at German schools should not presuppose that teachers are already versed in working in teams and reflecting critically together. In the initial stage of implementing LS, extra time for teachers to practice their reflection skills and establish norms for their collaborative work should be provided. In addition, the findings indicate that external facilitators and established protocols can play key roles in supporting LS teams in deepening their reflection. Policymakers and researchers interested in sustaining LS at German schools should thus consider how these resources, alongside addition resources identified by Jakobeit et al. (2021), can be provided to schools. For instance, future LS research could focus on identifying stakeholders that can provide external expertise on LS, as well as developing materials to ensure those stakeholders possess the necessary knowledge of LS. In general, LS materials and resources for teachers and school leaders (e.g., LS handbook adapted to the German context, reflection protocols, observation templates) could greatly scaffold implementation efforts in Germany. Importantly, these resources should be freely accessible to interested schools.

Turning to the second objective, this dissertation suggests several implications for research on LS in Germany. The current situation in Germany seems, in some ways, analogous to the US context in 2006, when Lewis an colleagues formulated their critical research needs for LS. Specifically, the current research base on LS in Germany is premature, as only few detailed documentations of LS and even fewer research studies on LS in the Germany context exist.

While LS is gaining momentum in Germany, it is not yet clear whether it can and will establish itself in the German school system or rather fade away and be replaced by another PD concept. If we were to conduct a randomized controlled trial on the efficacy of LS at German schools at this point in time, the results would likely resemble those of Lewis and Perry (2017) and Murphy et al. (2017). It would probably be difficult to establish high consistency within the groups, as the preconditions at German schools differ substantially (e.g., Jakobeit et al., 2021) and the ways in which teachers engage in LS would be hard to control (e.g., Lewis & Perry, 2017). A controlled evaluation would therefore likely fail to yield convincing evidence for the positive impact of LS and, as Maddux and Cumming (2004) remark, lead to disillusion with the idea of LS and its abandonment. Before putting LS to the test, this dissertation suggests that, similar to the US context in 2006, researchers and educators in Germany need to first enhance their understanding of LS and improve the ways in which LS is implemented in German school settings. DBR seems to be a promising way to do so. Looking forward, a systematic review of design principals identified as essential in DBR studies could greatly enhance our knowledge of effective LS in Germany and pave the way to more controlled research designs.

Unlike Lewis and colleagues in 2006, however, stakeholders in Germany have access to a vast international research base on LS that can greatly inform their research efforts. For instance, empirical research from around the world has produced a list of preconditions for LS, such as time, staff, support by school leaders, external expertise, and a clear understanding of the LS procedures (e.g., Bae et al., 2016; Lee & Madden, 2019; Lee & Tan, 2019; Mynott, 2019). These research findings can accelerate the implementation of LS in Germany and also lend weight to arguments put forward by researchers when seeking structural support from the government and policy makers (e.g., to provide teachers with additional time to engage in PD during their regular working hours). In order for LS to not be a fad in Germany, this dissertation therefore recommends to engage with the international discourse on LS and exchange knowledge with researchers and educators that are implementing LS in other national contexts. For example, several universities and teacher colleges in Austria have integrated LS into initial teacher education (e.g., Mewald, 2021, Mewald & Mürwald, 2021; Soukup-Altrichter et al., 2020). Prospective teachers become familiar with the concept of continuous and site-based PD already before they start teaching and enter their profession with the necessary knowledge to participate in, or even initiate, LS at their school. This

approach to integrate LS in the education system seems a promising avenue for Germany and should be explored in the future.

Finally, this dissertation proposed some recommendations for research on LS that also apply to research projects in German. As research on LS in Germany is still in its early stages, this is an ideal time to define terms, construct a common conceptualization, and adopt a shared language to communicate not only among researchers but also with educators and policymakers.

15. Conclusion

The objective of this dissertation was to explore the progress that has been made in the field of LS over the past two decades and to suggest directions for future research. The first chapter—Then—considered the beginnings of the research field and discussed the critical research needs (Lewis et al., 2006) which motivated and shaped early research efforts. The second chapter—Now—investigated to what extent these research needs have been explored and advanced over time. The analysis showed that, while LS has emerged as a "global phenomenon" (Hadfield & Jopling, 2016), it still sometimes fades away in national contexts and remains underdescribed and insufficiently understood in the research literature. Two scoping reviews of the LS literature indicated that progress on the research needs identified by Lewis et al. (2006) is hindered by several limitations. These are the frequent lack of comparable and replicable descriptions of the LS intervention in publications, the incoherent use or lack of use of theoretical frameworks to explain teacher learning through LS, the inconsistent use of terminology and concepts, and the lack of scientific rigor in research studies and of established ways or tools to measure the effectiveness of LS.

These limitations were addressed in three research studies. Study 1, the mixed-method study, addressed the theorization of the reflection stage and documented how four LS teams at German schools reflected together. Study 2, the systematic review, assessed how transparent publications are in describing how LS teams observed and reflected together, and whether theoretical underpinnings are made explicit. Study 3 proposed a conceptual model of LS in an effort to provide a shared reference point to the field. Together, the research studies generated several insights into the extent and the source of the problems that hinder progress in the field of LS.

The third part of this dissertation—Looking Forward—updated the research needs by Lewis et al. (2006) to reflect the current conditions and needs of the field and recommended several practices and strategies that can help advance these research needs in the future.

16. References

Aguinis, H., Ramani, R. S., & Alabduljader, N. (2018). What You See Is What You Get? Enhancing Methodological Transparency in Management Research. *Academy of Management Annals*, *12*(1), 83–110. https://doi.org/10.5465/annals.2016.0011

Akiba, M., Murata, A., Howard, C. C., & Wilkinson, B. (2019). Lesson study design features for supporting collaborative teacher learning. *Teaching and Teacher Education*, *77*, 352–365. https://doi.org/10.1016/j.tate.2018.10.012

Anderson, T., & Shattuck, J. (2012). Design-Based Research: A Decade of Progress in Education Research? *Educational Researcher*, *41*(1), 16–25. https://doi.org/10.3102/0013189X11428813

Bae, L. C., Hayes, K. N., Seitz, J., O'Connor, D., & DiStefano, R. (2016). A coding tool for examining the substance of teacher professional learning and change with example cases from middle school science lesson study. *Teaching and Teacher Education*, *60*(2016), 164–178. https://doi.org/10.1016/j.tate.2016.08.016

Bandura, A. (1969). Social-learning theory of identificatory processes. In D. Goslin (Ed.), *Handbook of socialization theory and research* (pp. 213–262). Rand McNally & Company. https://tiny.link.bandura

Baumfield, V., Bethel, A., Boyle, C., Katene, W., Knowler, H., Koutsouris, G., & Norwich, B. (2022). How lesson study is used in initial teacher education: An international review of literature. *Teacher Development*, *26*(3), 356–372. https://doi.org/10.1080/13664530.2022.2063937

Bjuland, R., & Mosvold, R. (2015). Lesson study in teacher education: Learning from a challenging case. *Teaching and Teacher Education*, *52*(2015), 83–90. https://doi.org/10.1016/j.tate.2015.09.005

BMBF, & KMK. (2016). *Gemeinsame Initiative von Bund und Ländern zur Förderung leistungsstarker und potenziell besonders leistungsfähiger Schülerinnen und Schüler. [Joint initiative by the federation and federal states for the support of gifted and potentially gifted pupils]*. https://www.bmbf.de/de/leistung-macht-schule-3641.html

Brown, C., Poortman, C., Gray, H., Groß Ophoff, J., & Warf, M. (2021). Facilitating collaborative reflective inquiry amongst teachers: What do we currently know? *International Journal of Educational Research*, *105*(101695), 1–11. https://doi.org/10.1016/j.ijer.2020.101695

Brown, C., Taylor, C., & Ponambalum, L. (2016). Using design-based research to improve the lesson study approach to professional development in Camden (London). *London Review of Education*, *14*(2), 4–24. https://doi.org/10.18546/LRE.14.2.02

Bryk, A. S. (2015). 2014 AERA Distinguished lecture: Accelerating how we learn to improve. *Educational Researcher*, *44*(9), 467–477. https://doi.org/10.3102/0013189X15621543

Callahan, C. (2019). Middle school geography teachers' professional development centered around historical photographs. *Journal of Social Studies Research*, *43*(4), 375–388. https://doi.org/10.1016/j.jssr.2018.11.003

Century, J., & Cassata, A. (2016). Implementation research: Finding common ground on what, how, why, where, and who. *Review of Research in Education*, *40*(1), 169–215. https://doi.org/10.3102/0091732X16665332

Cheung, W. M., & 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*, *3*(2), 137–149. https://doi.org/10.1108/IJLLS-05-2013-0024

da Ponte, J. P. (2017). Lesson studies in initial mathematics teacher education. *International Journal for Lesson and Learning Studies*, *6*(2), 169–181. https://doi.org/10.1108/IJLLS-08-2016-0021

Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective Teacher Professional Development*. Learning Policy Institute. https://learningpolicyinstitute.org/product/teacher-prof-dev.

Davidoff, F., Dixon-Woods, M., Leviton, L., & Michie, S. (2015). Demystifying theory and its use in improvement. *BMJ Quality & Safety*, *24*(3), 228–238. https://doi.org/10.1136/bmjqs-2014-003627

Elliott, J. (2019). Quality criteria for lesson and learning studies as forms of action research. *International Journal for Lesson and Learning Studies*, *9*(1), 11–17. https://doi.org/10.1108/IJLLS-02-2019-0018

Færøyvik Karlsen, A. M. (2019). Investigating teacher learning in Lesson Study: The important link between reported observations and change of plans. *Professional Development in Education*, *48*(1), 53-69. https://doi.org/10.1080/19415257.2019.1685564

Festinger, L. (1962). Cognitive Dissonance. *Scientific American*, *207*(4), 93–106.

Guskey, T. R. (2017). Where Do You Want to Get To? Effective Professional Learning Begins with a Clear Destination in Mind. *Educational School and Counseling Psychology Faculty Publications*, *38*(2), 32-37. https://uknowledge.uky.edu/edp_facpub/27

Guskey, T. R., & Yoon, K. S. (2009). What works in professional development? *Phi Delta Kappan*, *90*(7), 495–500. https://doi.org/10.1177%2F003172170909000709

Hanfstingl, B., Rauch, F., & Zehetmeier, S. (2019). Lesson study, learning study and action research: Are there more differences than a discussion about terms and schools? *Educational Action Research*, *27*(4), 455–459. https://doi.org/10.1080/09650792.2019.1652450

Harmon-Jones, E., & Mills, J. (2019). An introduction to cognitive dissonance theory and an overview of current perspectives on the theory. In E. Harmon-Jones (Ed.), *Cognitive dissonance: Reexamining a pivotal theory in psychology* (2nd ed., pp. 3–24). American Psychological Association. https://doi.org/10.1037/0000135-001

Huang, R., & Shimizu, Y. (2016). Improving teaching, developing teachers and teacher educators, and linking theory and practice through lesson study in mathematics: An international perspective. *ZDM*, *48*(4), 393–409. https://doi.org/10.1007/s11858-016-0795-7

Jaakkola, E. (2020). Designing conceptual articles: Four approaches. *AMS Review*, *10*(1–2), 18–26. https://doi.org/10.1007/s13162-020-00161-0

Jakobeit, M., Westphal, A., Jurczok, A., Kager, K., & Vock, M. (2021). Welche schulischen Rahmenbedingungen sind für die Kooperation von Lehrkräften in Lesson-Study-Prozessen förderlich? Eine Clusteranalyse von Grundschulen im Rahmen des Forschungs- und Schulentwicklungsprojekts Leistung macht Schule (LemaS) [What educational conditions are beneficial for teachers' collaboration in lesson-study projects? In the context of the research and school development project Leistung macht Schule, a cluster analysis of primary schools was conducted]. *Journal Für Schulentwicklung*, *2*(21), 17–26. https://www.mathtransformations.com/_files/ugd/e8bafa_519d47a1ee3544a3b6627006488fb271.pdf#page=18

Kager, K., Kalinowski, E., Jurczok, A., & Vock, M. (2021). *A systematic review of methodological transparency in lesson study: How do we report on the observation and reflection stages?* OSF. https://doi.org/10.17605/OSF.IO/5NXGY

Karlsen, A. M. F., & Helgevold, N. (2019). Lesson Study: Analytic stance and depth of noticing in post-lesson discussions. *International Journal for Lesson and Learning Studies*, *8*(4), 290–304. https://doi.org/10.1108/IJLLS-04-2019-0034

Knoblauch, R. (2019). Beobachtung und Dokumentation von Lernaktivitäten in der Lesson Study [Observation and documentation of learning activity in lesson study]. In C. Mewald & E. Rauscher (Eds.), *Lesson Study. Das Handbuch für kollaborative Unterrichtsentwicklung und Lernforschung* [*Lesson study. The handbook for collaborative lesson development and learning research*] (pp. 49-76). Studienverlag.

Korthagen, F. (2016). Inconvenient truths about teacher learning: Towards professional development 3.0. *Teachers and Teaching*, *23*(2), 387–405. https://doi.org/10.1080/13540602.2016.1211523

Korthagen, F. A. J. (1985). Reflective teaching and preservice teacher education in the Netherlands. *Journal of Teacher Education*, *36*(5), 11–15. https://doi.org/10.1177/002248718503600502

Korthagen, F. A. J. (2010). Situated learning theory and the pedagogy of teacher education: Towards an integrative view of teacher behavior and teacher learning. *Teaching and Teacher Education*, *26*(1), 98–106. https://doi.org/10.1016/j.tate.2009.05.001

Larssen, D. L. S., Cajkler, W., Mosvold, R., Bjuland, R., Helgevold, N., Fauskanger, J., Wood, P., Baldry, F., Jakobsen, A., Bugge, H. E., Næsheim-Bjørkvik, G., & Norton, J. (2018). A literature review of lesson study in initial teacher education: Perspectives about learning and observation. *International Journal for Lesson and Learning Studies*, *7*(1), 8–22. https://doi.org/10.1108/IJLLS-06-2017-0030

Lee, V., & Madden, M. (2019). "We're in this together": Principals and teachers as partners and learners in lesson study. *NASSP Bulletin*, *103*(1), 51–64. https://doi.org/10.1177/0192636519826717

Lee, L. H. J., & Tan, S. C. (2020). Teacher learning in Lesson Study: Affordances, disturbances, contradictions, and implications. *Teaching and Teacher Education*, *89*(2020), 102986. https://doi.org/10.1016/j.tate.2019.102986

Lewis, C., & Perry, R. (2017). Lesson study to scale up research-based knowledge: A randomized, controlled trial of fractions learning. *Journal for Research in Mathematics Education*, *48*(3), 261–299. https://doi.org/10.5951/jresematheduc.48.3.0261

Lewis, C., Perry, R., & Murata, A. (2006). How should research contribute to instructional improvement? The case of lesson study. *Educational Researcher*, *35*(3), 3–14. https://doi.org/10.3102/0013189X035003003

Lipowsky, F., & Rzejak, D. (2015). Key features of effective professional development programmes for teachers. *Ricercazione*, *7*(2), 27–51.

Maddux, C., & Cummings, R. (2004). Fad, Fashion, and the Weak Role of Theory and Research in Information Technology in Education. *Journal of Technology and Teacher Education*, *12*(4), 511–533. https://link.gale.com/apps/doc/A126331245/AONE?u=anon~c51e6067&sid=googleScholar&xid=0dd8bddf

Mann, S., & Walsh, S. (2013). RP or 'RIP': A critical perspective on reflective practice. *Applied Linguistics Review*, *4*(2), 291–315. https://doi.org/10.1515/applirev-2013-0013

Mercer, N. (2000). *Words and Minds: How We Use Language to Think Together.* Routledge.

Mewald, C. (2021). Lesson study in vocational education and training: The status quo in four European countries. *R&E-SOURCE*, *16*. https://doi.org/10.53349/resource.2021.i16.a998

Mewald, C., & Mürwald-Scheifinger, E. (2019). Lesson study in teacher development: A paradigm shift from a culture of receiving to a culture of acting and reflecting. *European Journal of Education*, *54*(2), 218–232. https://doi.org/10.1111/ejed.12335

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & and the PRISMA Group. (2009). Reprint—Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Physical Therapy*, *89*(9), 873–880. https://doi.org/10.1093/ptj/89.9.873

Moravcsik, A. (2020). *Transparency in qualitative research*. SAGE Publications Ltd. https://doi.org/10.4135/9781526421036

Murphy, R., Weinhardt, F., Wyness, G., & Rolfe, H. (2017). Lesson study: Evaluation report and executive summary. *Education Endowment Foundation*. https://doi.org/10.1016/0273-2297(83)90007-2

Mynott, J. P. (2019). Lesson study outcomes: A theoretical model. *International Journal for Lesson and Learning Studies*, *8*(2), 117–134. https://doi.org/10.1108/IJLLS-08-2018-0057

Mynott, J. P., & O'Reilly, S. E. M. (2022). Establishing a lesson study collaboration matrix. *International Journal for Lesson & Learning Studies*, *11*(3), 174–192. https://doi.org/10.1108/IJLLS-01-2022-0004

Parks, A. N. (2008). Messy learning: Preservice teachers' lesson-study conversations about mathematics and students. *Teaching and Teacher Education*, *24*(5), 1200–1216. https://doi.org/10.1016/j.tate.2007.04.003

Patton, M. (1990). *Qualitative evaluation and research methods* (2nd ed.). SAGE.

Pitzl, M. L. (2020). Tracing the emergence of situational multilingual practices in a BELF meeting: Micro-diachronic analysis and implications of corpus design 1. In K. Murata (Ed.), *ELF research methods and approaches to data and analyses: theoretical and methodological underpinnings* (pp. 97-125). Routledge.

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, *66*(2), 211–227. https://doi.org/10.1002/sce.3730660207.

Radovic, D., Black, L., Williams, J., & Salas, C. E. (2018). Towards conceptual coherence in the research on mathematics learner identity: A systematic review of the literature. *Educational Studies in Mathematics*, *99*(1), 21–42. https://doi.org/10.1007/s10649-018-9819-2

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

Rzejak, D. (2019). Zur Wirksamkeit von Lesson Study. In C. Mewald & E. Rauscher (Eds.), *Lesson Study. Das Handbuch für kollaborative Unterrichtsentwicklung und Lernforschung* (pp. 97–111). Studien Verlag.

Saito, E. (2012). Key issues of lesson study in Japan and the United States: A literature review. *Professional Development in Education*, *38*(5), 777–789. https://doi.org/10.1080/19415257.2012.668857

Seleznyov, S. (2019). Lesson study beyond Japan: Evaluating impact. *International Journal for Lesson and Learning Studies*, *8*(1), 2–18. https://doi.org/10.1108/IJLLS-09-2018-0061

Seleznyov, S., Goei, S. L., & Ehren, M. (2021). International policy borrowing and the case of Japanese Lesson Study: Culture and its impact on implementation and adaptation. *Professional Development in Education*, 1–15. https://doi.org/10.1080/19415257.2021.1973069

Slavit, D., & Nelson, T. H. (2010). Collaborative teacher inquiry as a tool for building theory on the development and use of rich mathematical tasks. *Journal of Mathematics Teacher Education*, *13*(3), 201–221. https://doi.org/10.1007/s10857-009-9136-x

Soukup-Altrichter, K., Steinmair, G., & Weber, C. (Eds.). (2020). *Lesson Studies in der Lehrerbildung: Gemeinschaftliche Planung und Evaluation von Unterricht im Lehramtsstudium* [Lesson study in teacher education: Collaborative planning and evaluation of instructions in initial teacher education]. Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-27748-2

Stigler, J. W., & Hiebert, J. (2016). Lesson study, improvement, and the importing of cultural routines. *ZDM: The International Journal on Mathematics Education*, *48*(4), 581–587. https://doi.org/10.1007/s11858-016-0787-7

Tinoca, L., Piedade, J., Santos, S., Pedro, A., & Gomes, S. (2022). Design-Based Research in the Educational Field: A Systematic Literature Review. *Education Sciences*, *12*(6), 1–19. https://doi.org/10.3390/educsci12060410

van Es, E. (2011). A framework for learning to notice student thinking. In M. G. Sherin, V. R. Jacobs, & R. A. Philipp (Eds.), *Mathematics Teacher Noticing. Seeing through Teachers' Eyes* (pp. 164–181). Routledge.

van Es, E., & Sherin, M. (2002). Learning to notice: Scaffolding new teachers' interpretations of classroom interactions. *Journal of Information Technology for Teacher Education*, *10*(4), 571–596. https://tinyurl.com/yc82twtp

Vygotsky, L. S. (1986). *Thought and language.* Massachusetts Institute of Technology.

Walsh, S., & Mann, S. (2015). Doing reflective practice: A data-led way forward: *ELT Journal*, *69*(4), 351–362. https://doi.org/10.1093/elt/ccv018

Willems, I., & van den Bossche, P. (2019). Lesson Study effectiveness for teachers' professional learning: A best evidence synthesis. *International Journal for Lesson and Learning Studies*, *8*(4), 257–271. https://doi.org/10.1108/IJLLS-04-2019-0031

Yoshida, N., Matsuda, M., & Miyamoto, Y. (2021). The landscape of Lesson Study: A methodology for teachers' professional development and educational research. In J. Kim, N. Yoshida, S. Iwata, & H. Kawaguchi (Eds.), *Lesson Study-Based Teacher Education: The Potential of the Japanese Approach in Global Settings* (pp. 29–50). Routledge.

Yurkofsky, M. M., Peterson, A. J., Mehta, J. D., Horwitz-Willis, R., & Frumin, K. M. (2020). Research on Continuous Improvement: Exploring the Complexities of Managing Educational Change. *Review of Research in Education*, *44*(1), 403–433. https://doi.org/10.3102/0091732X20907363