Teaching and Teacher Education 119 (2022) 103831



Contents lists available at ScienceDirect

Teaching and Teacher Education

journal homepage: www.elsevier.com/locate/tate



Research paper

Understanding teacher learning in lesson study through a cultural–historical activity theory lens



Iris Uffen [a, *], Siebrich de Vries [a], Sui Lin Goei [b, c], Klaas van Veen [a], Nellie Verhoef [d]

[a] University of Groningen, Teacher Education, Grote Kruisstraat 2/1, 9712 TS, Groningen, the Netherlands
[b] Windesheim University of Applied Sciences, Department of Human Movement and Education, PO Box 10090, Campus 2-6, 8000 GB, Zwolle, the Netherlands
[c] VU Amsterdam, Department of Teacher Education LEARN Research Institute, PO Box 7057, De Boelelaan 1105, 1007 MB Van der Boechorststraat 1, 1081 BT, Amsterdam, the Netherlands
[d] University of Twente, Educational Science, De Zul 10, 7522, NJ, Enschede, the Netherlands

ARTICLE INFO

Article history:
Received 27 September 2021
Received in revised form
9 June 2022
Accepted 13 July 2022
Available online 22 August 2022

Keywords:
Teacher learning
Teachers' professional development
Lesson study
Cultural–historical activity theory

ABSTRACT

Although lesson study (LS) is widely considered a promising approach to teachers' professional development, the variation in teacher learning in LS is not yet well understood. Using a cultural–historical activity theory lens, we identify aspects of LS as a learning activity of significance for teacher learning. Using mixed-methods analysis of questionnaire and interview data drawn from 17 secondary school teachers, we find multiple variables of relevance in relation to teacher learning. This study clearly shows that participants' understanding of and attitude toward LS are at least as important as how LS is conducted.

© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

In the Netherlands, lesson study (LS) is a relatively new professional development approach for teachers that originated in Japan and is spreading across the world (Stigler & Hiebert, 2016). It is characterized by teachers' collaborative and inquiry learning, a strong alignment with teachers' own teaching practice, and a combined focus on subject matter content and student learning. These features are strongly associated with effective professional development (Dudley et al., 2019; Lewis et al., 2009; Van Driel et al., 2012). During LS, a small team of teachers formulates a shared research question about a teaching or learning problem they have experienced during their daily practice. To answer this question, they design and teach a "research lesson" and collect data on student learning, such as observation or interview data. The LS team discusses and reflects on the collected data and formulates explicitly what they have learned as teachers (Dudley, 2015; Fernandez et al., 2016; Fujii, 2016).

An advantage of LS is its inherent capacity to allow teachers to adapt and interpret it within their own specific context (Fujii, 2016; Huang & Shimizu, 2016; Wolthuis et al., 2020). This approach is aimed to provide a learning setting in which teachers purposefully translate learning goals to their students' concrete behaviors that indicate that they are achieving the learning goal. In other words, participating in LS develops teachers' capacity to recognize and understand learning processes of their own students to allow for improved teaching (Lee, 2015; Lewis et al., 2019; Schipper, 2019). By keeping the focus on teachers' own practices, LS touches on problems that teachers actually experience during teaching. Furthermore, the collaborative setting allows teachers to join forces and share their previously acquired knowledge and skills. In addition, LS offers the opportunity to examine beliefs that might easily stay unexamined in an individual learning context (Lee, 2015).

Although LS is a promising approach to teacher learning, empirical evidence shows that in practice, not every teacher benefits from participating in LS and outcomes of teacher learning differ among teachers, even within the same team (De Vries & Roorda, 2019; Mynott, 2019; Skott & Møller, 2017). These varying results raise questions about how teachers experience the LS process related to outcomes of teacher learning. In this study, we therefore view the experienced differences in the extent to which

* Corresponding author.
E-mail addresses: i.uffen@rug.nl (I. Uffen), s.de.vries@rug.nl (S. de Vries).

https://doi.org/10.1016/j.tate.2022.103831
0742-051X/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

teachers learn from LS as an opportunity to better understand the process of teacher learning through LS by comparing teachers who perceived to have learned more with those who perceived to have learned less from the approach.

In our study, we apply a cultural–historical activity theory (CHAT) lens. The CHAT is a learning theory that emphasizes that learning is shaped by a collaborative process guided by a shared object of learning and is influenced by participants' intrapersonal experiences (Engeström, 1987; Engeström & Sannino, 2010), which makes the CHAT an interesting theory to LS.

Drawing on questionnaire and interview data in a small-scale (N = 17) mixed-methods study, we investigate through a CHAT lens how perceptions of teachers who perceived to have learned more from LS differ from those who perceived to have learned less. Our main research question is: what influences teacher learning in LS from a CHAT perspective?

2. Theoretical background

2.1. The CHAT

The CHAT provides an in-depth and dynamic analysis of learning activities that take into account interrelations among multiple components (Engeström, 1987; Roth et al., 2009). The theory posits that an interplay of seven components and four connecting aspects mediate collaborative learning toward a shared object of learning, as depicted in the triangular model in Fig. 1 (based on Engeström, 1987). Explained from an LS standpoint, the top of the triangle represents three interacting components: (1) an individual LS participant, (2) an object of learning shared by all LS participants about their own students' learning and thinking, and (3) LS tools, for example a research lesson plan, that mediate LS participants' learning about the object of learning. The actions LS participants take ultimately lead to (4) the outcome of teacher learning in terms of developed cognitions and modified behavior (Bakkenes et al., 2010; Desforges, 1995). The lower part of the triangle represents the components related to collaboration. How the actions take shape is affected by (5) formal and informal social rules and rules of conduct within the LS team and (6) the understanding of the rules, guidelines, and procedures by individuals and how they are shared among the LS participants. Last, the (7) division of labor refers to the working arrangements made during the LS process, how roles are divided among the LS team, and how results are shared with other members of the community (e.g., school administration, parents of students) (Engeström & Sannino, 2010; Lee & Tan, 2020; Wei, 2019).

Fig. 1 shows that between the components are several relationships, which Engeström (1987) refers to as dominant aspects of human learning activities: production, exchange, distribution, and consumption (Engeström, 1987; Sannino, 2011). These aspects are to some extent open to interpretation because of the dynamic nature of the theory (Engeström, 1987), but they nevertheless provide insights into teacher learning from LS. The following subsections provide an overview of these four relationships, as well as teacher learning and its outcomes.

2.1.1. Production

Production refers to the actions taken during a learning activity that pertain to the direct production of a learning outcome in terms of psychical products, developed behavior, or "products of mind" such as thinking patterns, knowledge, and skills (Engeström, 1987). In the context of LS, all these taken actions together form the actual LS.

Before the start of LS, the group must determine a set of preconditions (facilitation of, e.g., enough time, space, and access to useful resources). These preconditions are critical for professional development programs in general (Van Veen et al., 2010) – and therefore also for LS – to continue and have the potential to provide a qualitatively good learning process. Although the preconditions determine the actual existence of LS, they indicate little about the performance of LS. Wolthuis et al. (2020) highlight the importance of how the "general script" of LS is conducted. Because LS is teacher led and school based (Fernandez & Yoshida, 2004; Takahashi & Yoshida, 2004), it is often locally adapted – thus, the wide variety of LS variations evident in practice (Fujii, 2016; Huang & Shimizu, 2016; Wolthuis et al., 2020). This variation stresses the importance of examining teacher learning in the context of LS while also researching how teachers actually conduct LS.

In the Netherlands, schools often work with an LS variant elaborated by De Vries et al. (2016), who based their work on an American adaptation by Stepanek et al. (2007) and Dudley's (2011) UK version, which involves using case students. Case students typically represent types of students who might have distinctive learning needs regarding the same lesson goal. The general script

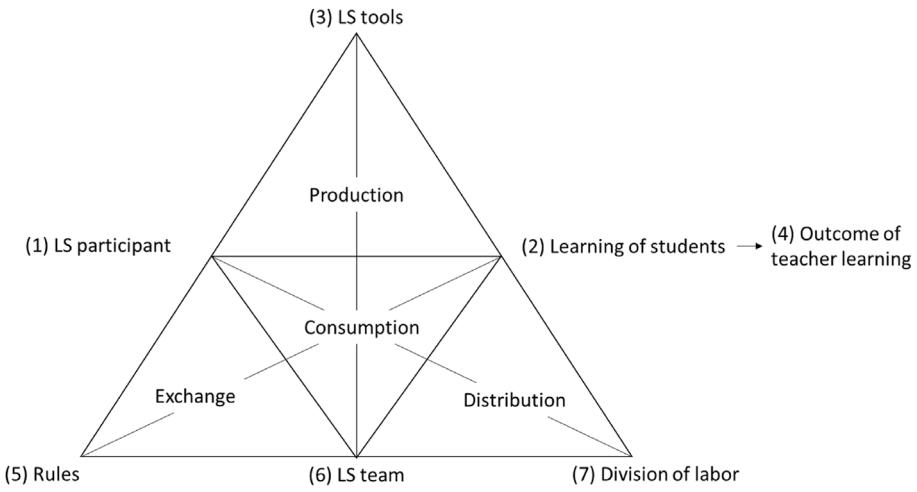


Fig. 1. Activity system model applied to lesson study.
Note. Adapted from *Learning by expanding. An activity-theoretical approach to developmental research* (p. 94), by Y. Engeström, 1987, Orienta-Konsultit. Copyright 1987 by Yrjö Engeström. Adapted with permission.

provided by De Vries et al. (2016) contains an LS cycle with six phases (depicted in Fig. 2).

During phase 1, the foundation (i.e., how LS participants will communicate and what their object of learning will be during the LS cycle) is set. Team members discuss how to collaborate, and they translate their object of learning into a shared research question in which they further specify what they want to learn about their object of learning. The research question sets the focus on student learning, which is considered the main goal of LS (Lee Bae et al., 2016; Choy et al., 2017).

In phase 2, the participants develop a research lesson that helps them address the research question. In this phase, the importance of a suitable research question emerges. Teachers select case students and predict their learning in preparation of the research lesson and the data collection (e.g., observation, interview data). Predicting student learning helps LS teams keep focus and alignment between the research lesson and the research question (Choy et al., 2017). Team members should not only share their own experiences and expertise but also study (lesson) materials and literature or consult an expert or knowledgeable others to deepen their knowledge (Takahashi, 2014). This access to external expertise can be used throughout the entire LS cycle.

Phases 3–5 are focused on teaching, re-teaching, and discussing the research lessons. During phase 3, one teacher deploys the research lesson, and the other LS team members observe and collect data on student learning. During phase 4, the participants conduct a post-lesson discussion in which they discuss and interpret their collected data in terms of student learning and consequences for their own learning as teachers. During phase 5, the team deploys a revised research lesson, and the members repeat phase 4.

Last, during phase 6 the participants reflect on the entire LS process and lessons learned and share their gained insights with the wider school staff. From a CHAT perspective, reflection and sharing results is an important part of the learning process, as participants together concretize what they have learned (Van Oers, 2009).

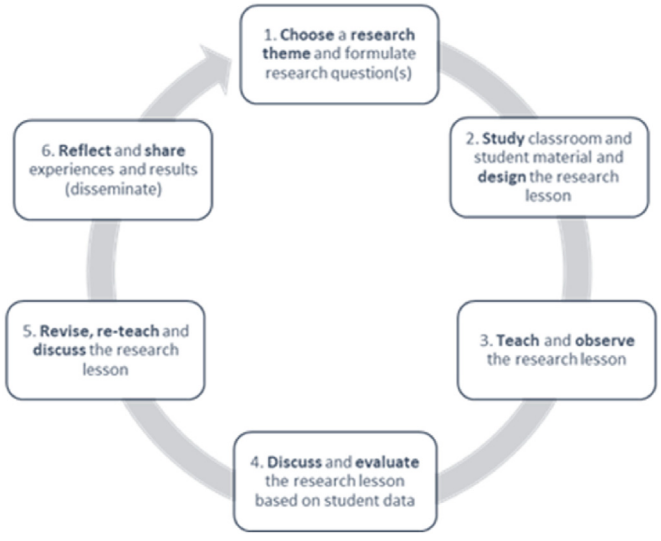


Fig. 2. Dutch lesson study cycle.
Note. Based on De Vries et al. (2016), from Teacher learning through Lesson Study. An examination of Lesson Study in relation to adaptive teaching, competence, teacher self-efficacy, and the school context (p. 20), by T. M. Schipper, 2019, Rijksuniversiteit Groningen. (https://doi.org/10.33612/diss.98636764). Copyright 2019 by Tijmen Schipper. Reprinted with permission.

In summary, we distinguish the following characteristics of the LS cycle, which prior research indicates are important for understanding how LS contributes to teacher learning: (1) preconditions of LS, (2) which phases the team conducted, (3) research question, (4) the focus on student learning, (5) use of external expertise, and (6) reflection and sharing results.

2.1.2. Exchange

Exchange refers to communication during a learning activity (Engeström, 1987). Several researchers suggest that the quality of LS conversations is essential for teachers to learn from one another (e.g., Dudley, 2013; Lee Bae et al., 2016; Vrikki et al., 2017). The challenge lies in creating a culture in which teachers feel free to express disagreement. Building on Mercer (1995), we refer to this "type of talk" as exploratory talk, during which LS participants engage in critical and constructive conversation by expressing doubts and digging deeper into the matter they discuss from various perspectives, which is associated with long-lasting personally meaningful learning (Fernandez et al., 2003; Parks, 2008). This stands in contrast to types of talk that are less critical in nature, such as cumulative talk in which teachers build on what the other had said without critically examining what is said, or qualifying and disqualifying talk in which teachers agree or disagree with one another without further exploring why they (dis) agree (Mercer, 1995). The latter can lead to their maintaining ineffective cognitions and teaching behavior and is ultimately counterproductive to development (Hargreaves, 2001; Nelson et al., 2010).

A space in which teachers have critical and constructive conversation does not exist by default, as interacting in such a way is often difficult (Nelson et al., 2010). LS participants need to feel secure enough with one another and their way of communicating to share their ideas and knowledge, and the facilitator has an important role in creating and supporting a trusting and open climate that establishes such security (De Vries & Uffen, 2021; Dudley, 2013; Salas et al., 2005). We refer to this feeling of safety as a necessary precondition of communication before explorative talk can occur. In summary, we divide exchange in two main concepts: (1) preconditions of the communication and (2) the extent in which teachers use exploratory talk.

2.1.3. Distribution

The third aspect of the learning activity pertains to the distribution of tasks during a learning activity (Engeström, 1987). In principle, LS is teacher led, which means that the participants themselves determine what and how they want to learn during LS (Fernandez & Yoshida, 2004). In the Netherlands, however, the LS team is commonly supported by a facilitator, especially teams that are new to LS. Several international studies point to the potential of a LS facilitator to contribute to the quality of the learning process (Amador & Weiland, 2015; Lee Bae et al., 2016; Lee & Tan, 2020). Facilitator-specific tasks include supporting and guiding the teachers through LS and optimizing the learning process for the participants by ensuring preconditions and the quality of the LS conversations (De Vries et al., 2016).

Sometimes, however, facilitators involve themselves more substantively in LS, depending on the team's needs. For example, an LS team that struggles to find an object of learning could benefit from concrete suggestions from a facilitator, or a team that struggles to draw meaning from the collected data during the research lesson could benefit from the facilitator's ideas and insights (Amador & Weiland, 2015; Lewis, 2016). In these situations, the facilitator may take on a more participatory role instead of the more typical supportive role. In summary, we distinguish how the participants perceive the extent to which (1) the facilitator conducted

specific facilitator tasks and (2) the facilitator took on a more participatory role during the LS conversations.

2.1.4. Consumption

Consumption lies at the center of the framework triangle and refers to whether the individual participant "consumes" the outcome of the learning activity's production process. During consumption, participants interpret the LS process and give meaning to the process in relation to themselves. In other words, consumption is the process by which participants decide – consciously or unconsciously – what is meaningful for them to learn (Engeström, 1987). Underlying consumption is emotional engagement, which is a good predictor of learning from an activity that relates to a perceived control, feeling of relevance and an overall positive attitude regarding the learning activity (Engeström & Sannino, 2010; Roth et al., 2009). Translated to the LS context, participants who feel emotionally engaged perceive control over how to perform LS, understand how the LS process leads to teacher learning, perceive the object of learning as relevant for themselves, and have an overall positive attitude toward LS, while being influenced by colleagues' attitudes toward LS. In summary, we distinguish perceived control over LS – in particular, the understanding of LS as a subvariable of perceived control – the perceived relevance of the object of learning, and the LS team members' attitudes toward LS.

2.1.5. Teacher learning

The focus of any learning activity is learning and development (Engeström, 1987). In LS, participants develop patterns of thinking and actions in their roles as teachers (Mynott, 2019; Takahashi & McDougal, 2016; Wei, 2019). From a CHAT perspective, LS can lead to expansive learning in terms of development of innovative new teaching and learning practices, often unexpected beforehand. Even smaller innovative activities might lead to more isolated but still personally meaningful inquiry-based learning, in which participants experiment with various ideas in their own teaching practice to challenge some of their cognitions and behaviors (Miedema & Stam, 2008). As learning can differ among participants and sometimes be unanticipated, we choose to view teacher learning through a broad lens by defining teacher learning as any outcome of changed cognitions and behavior (Bakkenes et al., 2010; Desforges, 1995; Desimone, 2009). In addition, because LS is inquiry based we have good reason to expect that LS participants will develop cognitions and behavior related to pedagogy (Lewis et al., 2009) and/or develop research literacy skills. In summary, we distinguish outcomes of teacher learning in terms of (1) any outcome of changed cognitions and behavior, as well as more specific changes in terms of (2) pedagogy and (3) research literacy skills.

2.2. Research questions

As noted previously, our main research question is: what influences teacher learning in LS from a CHAT perspective? Given the theoretical background, the following subquestions arise:

1. How does production influence teacher learning?
2. How does exchange influence teacher learning?
3. How does distribution influence teacher learning?
4. How does consumption influence teacher learning?
5. What are the outcomes of teacher learning from LS?

3. Method

3.1. Research design

With this study, we aim to gain insight into variables that can influence teacher learning from LS, as well as explore more contextual and in-depth insights into how and why these variables relate to teacher learning. For this reason, we chose to use a mixed-method design consisting of questionnaire and interview data. The use of a questionnaire allows for standardized testing and scaling the given answers between LS participants who perceived to have learned more or less from LS, which is mainly useful in gaining insight into which variables are relevant in relating teacher learning from LS.

The interview data are intended to show an explanatory sequential design by providing more in-depth and contextual information to explore how and why specific variables are relevant for teacher learning from LS. The use of both sources of data support the credibility of the findings (Devers, 1999), which adds value for this study, considering its small scale.

3.2. Research team

The research team consisted of five educational researchers who were jointly involved in the development of the questionnaire and the interview outline. The first author gathered the data in collaboration with the participating schools' LS coordinator.

3.3. Participants and procedure

We collected data at one secondary school in which all teachers (N = 20) participated in LS. The teachers were divided into four LS teams supported by three internal facilitators and one external facilitator from an education consultancy not related to the research institute of the authors.

The LS coordinator of the school sent all teachers an information brochure about the research project via email. They were asked to reply within one week if they wanted to participate and were given the opportunity to contact the first author by email or phone if they had questions. After a week, the LS coordinator provided the willing teachers with a link to the questionnaire by email; 17 teachers filled out the questionnaire completely, 7 of whom were open to an in-depth interview. Table 1 gives a general description of participants' characteristics. We do not provide characteristics of the interview participants; considering the small sample size, identification is highly probable.

All participants provided active informed consent before filling out the questionnaire, the only way they could enter the questionnaire. Of the 20 participants, 17 teachers filled out the questionnaire completely. We obtained ethical clearance to conduct the study from the ethical board of the teacher education department of the University of Groningen [TED-1920-S-0015].

3.4. Data collection

3.4.1. Quantitative data co llection

We collected quantitative data via a Qualtrics questionnaire. Except when an alternative source is indicated, the scales are based on Van Harskamp (2018), who developed and validated a preliminary version of this questionnaire.

The questionnaire consisted of 79 questions divided among 14 scales, one categorical question, and one open-ended question. All the scales measured items on a 5-point Likert scale (1 = strongly disagree, and 5 = strongly agree) and had sufficient internal consistency, as evidenced by Cronbach's alpha of .70 or higher (ranging

from $\alpha = 0.703$ to $\alpha = 0.946$). Table 2 presents example items and Cronbach's alpha per scale.

The questions and scales on production consisted of (1) one categorical question in which teachers reported the specific LS phases they conducted, (2) one open-ended question for them to identify their research question, and the four scales measuring (3) preconditions (i.e., time, space, and resources), (4) the focus on student learning, (5) the use of external expertise, and (6) reflection and sharing results. The scale on preconditions was slightly adjusted for this study due to low reliability in the study by Van Harskamp (2018). The last three scales were adjusted and expanded given the difference of context: Van Harskamp (2018) designed the scales to gain insight on the influence of the facilitator, and the items for this study are intended to collect information on what occurred during the LS cycle despite the influence of the facilitator (for example: "The LS facilitator stimulated the use of (scientific) literature" changed to "We read books or other publications to deepen our knowledge.").

The scales on exchange were divided into (1) the preconditions of the communication and (2) the extent of use of exploratory talk between the team members. The first scale was originally developed by De Vries et al. (2017), who based their questions on Salas et al. (2005). We developed the second scale for this study, building mainly on Dudley (2013) and Mercer (1995).

The scales on distribution were divided into (1) the facilitator's participatory role and (2) facilitator-specific tasks. The second scale was based on De Vries et al.'s (2016) description of typical tasks for facilitators. The scale on consumption consisted of (1) participants' feeling of control on LS, (2) participants' attitude toward LS, and (3) how participants perceived their colleagues' attitudes toward LS. The scale on perceived control was adapted from Ehren (2006) and adjusted to the LS context. Both the scales on perceived control and participants' attitude toward LS were slightly adjusted due to low reliability in the study by Van Harskamp (2018) and a following pilot study by the first author.

The scale on teacher learning is divided into (1) teacher learning in relation to pedagogy and (2) teacher learning in relation to research literacy skills. The items of the first scale and one item of the second scale are based on Lewis (2005) by De Vries et al. (2017) and validated within the Dutch context. We developed all other items for the second scale.

3.4.2. Qualitative data collection

We gathered the qualitative data using an in-depth interview in which we used vignettes and semi-structured questions. The goal of the vignettes was to gain in-depth insight into how LS participants perceived specific LS phases. Each vignette consisted of a fictional summary of an LS team meeting based on the LS cycle as discussed in Section 2.1.1. The purpose of the vignettes was to recall memories about the participants' own LS cycle and ideas on LS in general. An additional intention was to trigger participants to share their understanding of LS by presenting fictional cases containing some common misinterpretations about (how to conduct) aspects of LS. For example, a trigger point could be how to formulate a learning goal for the research lesson. The Appendix provides an example of a vignette.

The first author asked the participants individually to discuss two vignettes during an approximately hour-long interview (interviews ranged from 35 to 85 min). To avoid biases, the wording of the questions was taken into account (e.g., neutral formulation) and the interviewer told the interviewees there were no "good" or a "bad" answers and that answers that most closely reflected their own experience would fit the research best. Furthermore, all the vignettes shared the same question pattern. First, the participants related how they perceived what the fictional LS team did, what the LS team talked about, and the collaboration of the LS team during the specific LS phase covered in the vignette. Second, participants responded to the same questions but in the context of their own LS team. Third, they described what they as an LS team could have done differently. Fourth, they identified what they perceived the relevance of the specific phase to be to LS in general and for their specific LS cycle. Finally, they responded to two specific questions: one pertaining to the research question of their LS cycle and one about what they learned from LS.

The purpose of the first and second sets of questions was to obtain information specifically on production, exchange, and distribution; the third and fourth set of questions addressed consumption; and the last set focused on outcome of teacher learning. In practice, we received information on all four aspects of a learning activity and outcome of teacher learning throughout the interviews, depending on what the question triggered for the participant.

3.5. Data analysis

3.5.1. Quantitative data analysis

We analyzed the quantitative data in SPSS version 25.0 using descriptive and test statistics. We divided the participants into two groups using median split: teachers who indicated that they learned more (higher teacher learning; HTL, n = 8) and the group of participants on or below the median, who indicated they learned less (lower teacher learning: LTL, n = 9) than the group above the median. Considering the small sample size, we used an independent *t*-test with bootstrapping. This method allows estimating the

Table 1
Description of participants (N = 17).

Descriptive	Data
Gender	Male n = 8 (47%) Female n = 8 (47%) Unknown n = 1 (6%)
Age (in years)	M = 41.38/SD = 10.85/range 24–64 (missing n = 3)
Teaching experience (in years)	M = 14.82/SD = 9.82/range 3–38 (missing n = 0)
Teacher qualification	M.Ed. n = 4 (24%) B.Ed. n = 13 (76%)
Teaching subject	Creative subjects n = 2 Humanities n = 5 Languages n = 6 Math & science n = 5 Other n = 1
Past experience with LS	0 LS cycles n = 10 (59%) 1 LS cycle n = 7 (41%)

Table 2
Domains and scales of the questionnaire, examples of items, number of items, and Cronbach's alpha (if applicable).

Aspect learning activity	Domain or scale	Example of items	No. of items	Cronbach's α
Production	Conducted LS phases	"What LS phases did you participate in?"	1	d.n.a. – categorical
	Object of the research question	"What was the research object of the LS?"	2	d.n.a. – open ended
	Preconditions LS	"There was enough time."	7	.709
	Focus on student learning	"We made the learning and thinking of students visible during the research lesson."	5	.786
	Use of external expertise	"We studied lesson materials."	6	.776
	Reflection and sharing results	"We translated what we had learned to what it meant for our own teaching."	6	.817
Exchange	Preconditions of communication	"Team members trusted each other."	7	.876
	Exploratory talk	"We expressed our doubts when we thought something was not right."	4	.703
Distribution	Facilitator's participatory role	"The facilitator also took part in the discussion in terms of content."	7	.878
	Facilitator-specific tasks	"The facilitator stimulated a focus on the learning of students."	6	.725
Consumption	Perceived control LS	"I have enough research skills to conduct LS."	6	.877
	Participant's attitude	"LS contributes to my professional development."	6	.881
	Colleagues' attitudes	"My colleagues find LS inspiring."	6	.795
Teacher learning	Pedagogy	"Due to LS, my understanding of learning of students is improved."	6	.946
	Research literacy skills	"Due to LS, my skills in researching my own teaching practice are improved."	4	.933

significance of the found differences using a small sample size without the requirement of normality or equality of variance (Dwivedi et al., 2017). Table 3 describes the mean scores of the two groups. Questionnaire participants were equally distributed among the four LS teams: five teachers of one team and four teachers of the other three teams filled out the questionnaire.

3.5.2. Qualitative data analysis

Five HTL teachers and two LTL teachers participated in the interviews. We gave all interviewees pseudonyms with HTL or LTL added as appropriate. The HTL teachers were divided over three of the four LS teams (one team consisted of only LTL teachers) and the LTL teachers were divided over two teams.

After having the interviews transcribed verbatim, we analyzed the qualitative data in two steps. First, we developed a coding scheme based on the theoretical framework by analyzing the data and discussing the meaning of the codes within the research team. Second, we discussed the coded segments of the data to form a consensus of the meaning of the found results and to determine saturation.

After analyzing the data, we found that not all codes based on the theoretical framework applied to the interview data. For production, we divided preconditions of LS into a subvariable of "time" and external expertise into a subvariable of "use of literature." We did not code "conducted LS phases," "focus on student learning," and "reflection and sharing results" because either participants did not mention these things or their answers were too fragmented to move beyond arbitrary input. For consumption, we divided "perceived control over LS" into a subvariable of "understanding of LS," and we added "relevance of the object of learning." Exchange and distribution remained according to the questionnaire. Last, within teacher learning we divided "pedagogy" into the subvariables "gaining awareness and general insight," "development of the professional community and solidarity with colleagues," and "no teacher learning." We did not include "research literacy skills" in the coding scheme, because participants did not mention this variable. Table 4 presents the main codes of the scheme and how they relate to the questionnaire variables.

In addition to the research team's analysis, the first author performed a subject review by discussing the findings with the participating schools' principal and the team's facilitators. The meeting served the additional purpose of providing a LS evaluation for the school as a whole. When asked whether they recognized the findings from the quantitative and qualitative analyses, the facilitators confirmed that the presented findings fit their own experiences with the teams.

Table 3
Group division based on lower or higher teacher learning per individual teacher.

Scale	HTL (n = 8)		LTL (n = 9)		*p*-Value
	Mean	SD	Mean	SD	
Pedagogy	3.13	.36	1.76	.77	.005
Research literacy skills	3.25	.42	1.61	.69	.001

4. Findings

4.1. Production

4.1.1. Questionnaire data

All participants engaged in LS cycles in which all six phases were conducted. The questionnaire data (see Table 5) show that the HTL teachers scored significantly higher than the LTL teachers on the use of external expertise – which is still quite low for both HTL and LTL teacher, being the only variable with both means below 3.00 – and reflection and sharing results. With regard to preconditions of time, space, and access to resources as well as with regard to focus on student learning, the data do not show a significant difference between HTL and LTL teachers.

For the open-ended question (see Table 6) about what groups used as an object of learning, HTL teachers reported using peer feedback, both HTL and LTL teachers reported using ownership and intrinsic motivation of students and students as independent problem solvers, and LTL teachers reported using collaboration of students. We note that the use of peer feedback seems more specific and learning-task-related than the other objects of learning.

4.1.2. Interview data

The interview data provide additional insights into time (as a subvariable of preconditions), use of literature (as a subvariable of external expertise), and the perception of how the team translated the object of learning into a research question. We discuss each in turn.

First, regarding time, during the interviews, both HTL and LTL

Table 4
Overview questionnaire variables and additional interview codes.

Aspect learning activity	Questionnaire variable	Additional interview code
Production	Preconditions	*Subvariable*: time
	Conducted LS phases	–
	Research object	Research object
	Focus on student learning	–
	External expertise	*Subvariable*: use of literature
	Reflection and sharing results	–
Exchange	Preconditions of communication	Preconditions of communication
	Explorative talk	Explorative talk
Distribution	Facilitator's participatory role	Facilitator's participatory role
	Facilitator-specific tasks	Facilitator-specific tasks
Consumption	Perceived control LS	*Subvariable*: understanding of LS
	Participants' attitude	Attitude
	Colleagues' attitudes	Colleagues' attitudes
	–	Relevance object of learning
Teacher learning	Pedagogy	*Subvariable*: gaining awareness and general insights
	Research literacy skills	–
	–	Development community and solidarity
	–	No teacher learning

teachers recalled that they did not have enough time to conduct LS. Except for one HTL teacher, all participants' connected time with the quality of the LS cycle, such as lack of time to read literature in depth or to thoroughly prepare and develop a research lesson. For example, Chris-HTL said:

> Because I think for what should be an optimal LS, we were running out of time. So we really had to limit some phases here and there.

Second, regarding use of literature, HTL teachers except Chris-HTL and both LTL teachers mentioned that they made limited use of literature and did not read the literature in depth. The interview data provided no further insights into whether the HTL and LTL teachers experienced the lack of in-depth use of literature differently.

Third, regarding the object of learning, the LTL teachers explicitly stated that the translation of the object of learning into a research question was too vague and abstract. As a result, they mentioned that their research question and research lesson did not align. Evan-LTL explained:

> It was very difficult for us to formulate a research question relating to the subject of ownership. If that research question was not good, it is difficult to plan the [research] lessons and formulate a lesson goal. It just went wrong, I think.… How the lesson was conducted … if you want to answer the research question … that was far from each other.

Table 5
Mean scores on variables of production for HTL and LTL teachers.

Scale	HTL (n = 8)		LTL (n = 9)		*p*-Value
	Mean	SD	Mean	SD	
Preconditions	3.68	.37	3.30	.68	.186
Focus on student learning	3.88	.34	3.44	.56	.088
External expertise	2.80	.52	1.96	.81	.024
Reflection and sharing results	3.92	.45	3.00	.68	.005

Table 6
Research object distribution between HTL and LTL teachers.

Research object	HTL (n = 8)	LTL (n = 9)
Use of peer feedback	4	–
Ownership and intrinsic motivation	2	3
Independent problem solvers	2	2
Collaboration between students	–	4

4.2. Exchange

4.2.1. Questionnaire data

The questionnaire data (see Table 7) show no significant difference between HTL and LTL teachers on the preconditions of the communication. The data indicate that the HTL teachers perceived the use of exploratory talk as significantly greater than the LTL teachers.

4.2.2. Interview data

The interview data provide additional insights into preconditions of communications, as well as some limited additional insights into exploratory talk. We discuss each in turn.

First, both HTL and LTL teachers mentioned the already good preconditions of communication, which confirms the questionnaire results showing that HTL and LTL teachers did not experience the preconditions of communication differently. For example, Jesse-HTL referred to the equality of team members and active participation in LS, and Noah-HTL mentioned the good relationships:

> The way we interact with each other is actually very spontaneous and very direct. Those relationships are just very good, very open, and if there is something wrong, there is also the freedom to say that to each other. So, there are no repressed negative emotions and such.

Max-LTL referred more indirectly to the necessity of good preconditions when explaining why agreeing to codes of conduct is unnecessary:

> At our school that is absolutely unnecessary. That is no issue for us.

Second, the interview data show limited insights into the use of

Table 7
Mean scores on variables of exchange for HTL and LTL teachers.

Scale	HTL		LTL		*p*-Value
	(n = 8)		(n = 9)		
	Mean	SD	Mean	SD	
Preconditions communication	3.98	.14	3.70	.47	.144
Exploratory talk	3.72	.47	3.11	.55	.028

exploratory talk. Overall, HTL teachers sketched an ambiguous picture. Robin-HTL gave some indication of use of exploratory talk, saying:

> I found it striking that the conversation went in depth and got personal.

Both Bob-HTL and Noah-HTL, by contrast, highlighted the lack of exploratory talk, mentioning that the talk lacked in-depth interaction. For example, Bob-HTL said:

> Draw conclusions … we do that quite quickly.

Noah-HTL said:

> Gosh, we have taken it very lightly. I don't think we really got the depth of a good LS cycle out of it. I think we did it a little superficially.

The LTL teachers did not mention characteristics of exploratory talk or the lack thereof. In addition, one HTL teacher explicitly mentioned not knowing what to say about how the team members interacted with one another.

4.3. Distribution

4.3.1. Questionnaire data

A section of the questionnaire data (see Table 8) was focused on the facilitator's role and tasks. The results show that the HTL teachers perceived the facilitator as conducting facilitator-specific tasks significantly more than did the LTL teachers. However, the data reveal no significant difference between HTL and LTL teachers regarding the facilitator's participatory role.

4.3.2. Interview data

The interview data give limited additional insights into facilitator-specific tasks and no additional insights into the facilitator's participatory role. The interview data partially support the significance between the HTL and LTL teachers relating to the facilitator's specific tasks: four HTL teachers and one LTL teacher mentioned their facilitator's guidance to some extent, relating to conversational tactics such as asking questions and stimulating the use of literature. By contrast, Max-LTL mentioned the lack of facilitator guidance, expressing what he missed from the facilitator:

> I have got one thousand other things to do, so please take me by the hand and show me what is helpful because I want to, but you have to get me through. Because we are also with four colleagues who all want something, and we all want something that is useful for our [teaching] subject.

Table 8
Mean scores on variables of distribution for HTL and LTL teachers.

Scale	HTL		LTL		*p*-Value
	(n = 8)		(n = 9)		
	Mean	SD	Mean	SD	
Facilitator-specific tasks	4.02	.52	3.22	.71	.019
Facilitator's participatory role	3.70	.34	3.38	.88	.358

4.4. Consumption

4.4.1. Questionnaire data

The questionnaire data (see Table 9) show that the HTL teachers scored significantly higher on their attitude toward LS and also perceived their colleagues' attitudes toward LS as being significantly higher. With regard to the perceived control over LS, the data show no significant difference between the HTL and LTL teachers.

4.4.2. Interview data

The interview data give additional insights into LS and the understanding of LS as a learning activity as a subvariable of participants' perceived control over LS, participants' attitudes toward LS as well as their colleagues', and the perceived relevance of the object of learning. We discuss each of these topics in turn.

First, whereas the questionnaire data show no significant differences in perceived control over LS between HTL and LTL teachers, the interview data shed a different light on the subvariable "understanding of LS as a learning activity." For example, HTL teachers tended to explain how LS works, and LTL teachers tended to have questions about how LS works. The HTL teachers varied in their answers, as they addressed different LS phases. For example, Robin-HTL said, "*I believe you have to do the [research] lesson three times, in an optimal setting*"; Chris-HTL noted that during the research lesson, it is helpful to observe "*how [the students] worked in various ways [on their learning task]*"; Noah-HTL explained the importance of "*ask[ing] further question[s] [to the students] during the interview: what was the meaning of what happened during the lesson?*"; Bob-HTL emphasized the importance of "*checking whether our cause-effect reasoning is correct*"; and Jesse-HTL noted the importance of challenging one's own assumptions:

> It sounds bad, but I don't mean it [like that]. At what point do I cast pearls before swine? Which is frustrating for students, because they think, "I just don't get it at all," and frustrating for me, because I have expectations. I think, "I explained everything very well, now they will understand." But at some point you can end up having tunnel vision, in which it becomes difficult to view what you do [as a teacher]. I think LS can be super suitable for that.

In comparison, for example, Evan-LTL asked:

> I think it really didn't work out for us … also because we didn't have enough experience. What is LS actually?

Table 9
Mean scores on variables of consumption for HTL and LTL teachers.

Scale	HTL		LTL		*p*-Value
	(n = 8)		(n = 9)		
	Mean	SD	Mean	SD	
Perceived control LS self	4.08	.52	3.57	.71	.115
Participants' attitude	3.86	.39	2.70	.81	.009
Colleagues' attitudes	3.29	.38	2.69	.54	.018

And Max-LTL contemplated:

> In fact, I found myself in that whole process of LS experiencing that we actually had too little background. We started pretty quickly. I always had the question how can you only observe student behavior?

Another difference we found in the understanding of LS as a learning activity is that one HTL teacher – Noah-HTL – mentioned the influence of the context as an opportunity to learn, in contrast with the LTL teachers, who viewed the research lesson's context as a barrier to acquiring valuable data from the research lesson. Noah-HTL noted:

> The first [research lesson] was given the last hour of the day. The second [research lesson] was during the last hour of the week. You notice that the students are really looking forward [to] wrapping up the lesson. Those tiny details can leave your expectations unfulfilled.… On the other hand, this question [about student engagement] comes up every day, and every teacher approaches this in his own way and comes up with his own solutions – which only increases my package of possibilities.

As another example, Max-LTL said:

> It matters a lot where students are sitting … and if all students are there. We had a class of which four or five students were sick, and those are the present students. That has so much … that just determines your lesson.… Whether students are going to write something down has directly to do with who is sitting next to them.… To what extent can we measure objectively? Does it still have value?

Second, with regard to attitude toward LS, the interview data underpin the finding that HTL teachers have an overall more positive attitude toward LS than the LTL teachers. In more depth, the interview data show that HTL teachers' reasons for being positive are linked to explicitly valuing learning from the interaction between colleagues, in contrast with the LTL teachers, who reported not explicitly valuing the interaction between colleagues. For example, HTL teachers explained that they learn from different perspectives and ideas and observing how colleagues handle situations differently. Robin-HTL said:

> Every colleague has a different way of doing things and has a different understanding. That helps me to develop my own point of view. That doesn't mean that I necessarily agree with all the ideas of colleagues. Absolutely not. But even the ideas of which I think, "I would never do that" are helping me as well to take position. The conversation is really important to me.

Moreover, the interview data underpin the questionnaire data finding that the HTL teachers perceived the colleagues from their team as having overall more positive views toward LS or became more positive during the LS than the LTL teachers, who overall viewed the colleagues from their own team as having more negative attitudes toward LS. For example, Chris-HTL noted, "*We were all enthusiastic*," and Bob-HTL said:

> Some [colleagues] were not open to LS [in the beginning], but they did open to LS.

One LTL teacher, Evan-LTL, perceived his colleagues' positive as well as negative views but mentioned that he did not understand why some colleagues were positive about LS. Evan-LTL made a direct connection to learning from LS: "*I do not remember exactly what colleagues said they have learned, but I remember thinking: 'Is this all? Are you satisfied with so little?'*"

Third, regarding the relevance of the object of learning, the data show no difference between the HTL and LTL teachers. Both HTL and LTL teachers perceived the object of learning as relevant.

4.5. Outcomes of teacher learning

Teacher learning was the variable on which we divided the participants within the two groups of HTL and LTL (see Table 3, Section 3.5.1); therefore, we focus on the interview data for this topic, as they provide additional insights. Three HTL teachers mentioned insights into and awareness of pedagogy, and two HTL teachers mentioned that LS contributed to the development of a professional community and solidarity with colleagues. For example, Noah-HTL referred to learning relating to pedagogy by explaining that he learned to ask himself specific questions:

> How are they [the students] going to do that with each other? How are they going to react to each other? How are they going to motivate each other? How are they going to show commitment to the assignment they came up with together?

In relation to the development of a professional community and solidarity between colleagues, Bob-HTL mentioned:

> That means that you [visit each other's lessons] without the focus being on you. And that is also very good for the school. There is shyness in that. I think that is also a gain from LS … the indirect effect to make it safer [to visit each other's lessons].

Both LTL teachers expressed that they learned nothing from LS. For example, Evan-LTL said:

> We didn't really benefit from that either.… We also did not gain new insights.

5. Discussion and conclusion

Our aim was to gain insight into teacher learning from LS as a learning activity through a CHAT lens. The CHAT proves useful in operationalizing variables of relevance for teacher learning in LS and helps clarify that not only does LS by itself influence teacher learning, but LS participants' intrapersonal experiences, such as understanding of LS, also influence teacher learning to a great extent. Our findings specifically highlight the importance of the translation of the object of learning into a research question and its alignment with the research lesson, the use of external expertise, participants' understanding of LS, and participants' attitudes toward LS.

First, the *translation of the object of learning into a research question and its alignment with the research lesson* turned out to be crucial for the perceived relevance of the LS cycle in general. This finding is in line with Choy, Thomas, and Yoon (2007) and contrasts with our premise that the relevance of the object of learning relates to teacher learning (Roth, 2007); the perceived relevance of the object of learning on its own did not appear to be a distinguishing factor in relation to teacher learning. It could be argued that the object of learning maintains relevance when translated into a precise and focused research question, but it loses practical relevance if the translation into a research question is too vague and broad (as noted by LTL teachers), which leads to a loss of focus and alignment.

In practice, LS facilitator could support teams in identifying the focus of the research question – for example, by making sure that all of the participants know what is meant by the concept(s) in the research question, whether they want to focus on specific aspects of the used concept(s) for feasibility reasons, and how participants can evoke as well as recognize the (aspects) of the concept(s) in student learning during the research lesson. Future research could further explore how the alignment between the object of learning, research question, and research lesson can be strengthened to support teacher learning from LS.

Second, the *use of external expertise* stands out as an important influence on teacher learning. Although HTL teachers and LTL teachers both primarily used LS team members' expertise, the differences are significant between these groups, in line with a pilot study (Uffen, De Vries, Van Harskamp, & Goei, 2019, April 5-9) showing through multiple regression analysis that the use of external expertise is one of the stronger predictors for teacher learning and in line with the common recommendation that using external expertise (e.g., literature, a knowledgeable other) is a crucial feature to strengthen teacher learning through LS (Choy & Lee, 2021; Lewis & Perry, 2015; Takahashi, 2014). Although tapping external expertise is important for teacher learning, we find evidence that LS teams do not necessarily make use of it, in line with Choy and Lee's (2021, p. 39) finding that the study of materials for teaching (e.g. textbooks, literature) is an "often neglected aspect of lesson study" beyond Japan. In the Dutch context, Van den Boom-Muilenberg et al. (2021) suggest that the use of external expertise – especially literature – can be perceived as a complicating step in LS, and doing so may encounter resistance from participants. The participants of the study recommend omitting this step to "persuade" teachers to participate (p. 17). Other research suggests that even though some teachers might value the use of external expertise, they still do not tap external expertise of their own accord during LS professional development and find it difficult to use (Assen & Otting, 2022; Choy & Lee, 2021; Van Eekelen et al., 2006).

In practice, Choy and Lee (2021) point out that it is crucial that LS participants understand that the use of external expertise is key in the inquiry-oriented process of LS. External expertise enables essential encounters with new knowledge and perspectives (Koffeman & Snoek, 2018). The facilitator can stimulate the need for new knowledge and perspectives by asking inquiry-oriented questions for which external expertise is needed. A facilitator can further help select readable, high-quality literature, textbooks, and curriculum guides, and can support LS participants develop study skills. In addition, a facilitator can help search for other sources of external expertise, such as a knowledgeable other (Choy & Lee, 2021).

For future research, it would be fruitful to explore why participants do not value the use of external expertise – especially the use of literature – as essential, how participants can be motivated to use external expertise, and how various kinds of external expertise can be utilized to influence teacher learning from LS.

Third, an *understanding of LS* seems pivotal in experiencing teacher learning in the context of LS. Participants need to have their own understanding of how LS contributes to their learning, as this understanding facilitates ideas on how LS needs to be contributed; by contrast, not knowing how LS contributes to their learning evokes questions on how LS should be conducted. This finding is in line with Wolthuis et al. (2020), who show that participants' understanding of LS influences how they conduct it. Our study adds the further insight that understanding of LS influences not only how LS needs to be conducted but also how participants perceive the meaning of the conducted LS in retrospect. In practice, it is important that facilitators support participants in understanding what LS is and is not (De Vries & Uffen, 2021). During the LS cycle, the facilitator can point out explicitly why specific aspects are important for teacher learning during the cycle to ensure participants are consciously noticing learning points LS provides. For example, the facilitator can keep the group focused on student learning if participants start to shift toward a focus on teacher instruction or student behaviors unrelated to what they wanted to research.

Furthermore, this study suggests that participants' understanding of the research paradigm that fits LS has an influence on teacher learning. Beliefs that resonate with constructivism or critical theory (Lincoln et al., 2005) seem to relate to HTL, in the sense that participants view subjectivity as a learning opportunity. More positivistic beliefs about research (Lincoln et al., 2005) seem to relate to LTL, in which participants experience subjectivity as a barrier to learning from LS. Participants who strive for objectivity might strive for more quantitative research quality measures such as validity and reliability of the findings. Therefore, they might spend time trying to meet infeasible research quality measures: during LS, the data are collected within the varied context of education with a small sample size, and researchers have limited options to conduct controlled experiments, which makes it difficult to reach the requirements of validity and reliability. By contrast, participants who do not strive for objectivity might spend their time attempting to meet quality measures that relate to more qualitative approaches (e.g., credibility, authenticity) (Devers, 1999), which is more feasible with LS. These participants are probably more likely to spend their time on the essence of LS—that is, trying to understand how their students learn given the influence of the context.

In practice, a facilitator could support participants by making their varying views on research explicit and showing that one view contributes best to the aim of LS: learning to recognize and understand the learning of one's own students to allow for better and more informed teaching (Lee, 2015; Lewis, 2005; Schipper, 2019). We recommend that future research explore how beliefs about research relate to how LS participants conduct and interpret LS.

Fourth, we found that the *attitude toward LS* relates to teacher learning. More specifically, participants who explicitly mention valuing interaction with their colleagues experience HTL, which might indicate that some LS participants are focused more on learning from colleagues than others and therefore are more likely to learn from a collaborative activity such as LS (Dudley, 2013; Lee Bae et al., 2016).

In practice, schools could specifically focus on increasing experience with collaborative professional development. In the Netherlands, teachers are not used to learning collaboratively, and the inquiry-oriented conversations inherent in collaboration can conflict with already established conversational routines or even leave teachers feeling incapable or insecure (Arminen, 2005; De Vries & Uffen, 2021; Nelson et al., 2010). Therefore, schools should provide opportunities for safe conversations in which participants can explore their own ideas and assumptions freely and openly (Mercer, 1995) – what we refer to as the preconditions of communication. For example, all team members could form an explicit agreement that the team welcomes all ideas and explicitly values contradicting ideas that are a good starting point to broaden one's own perspectives and enhance learning (as stated by HTL teachers).

Furthermore, the finding that LTL teachers did not explicitly value interaction with colleagues also seems to relate to LTL teachers' lack of understanding of LS. The lack of understanding triggers questions related to LS, which might result in decreased time devoted to learning from their colleagues' perspectives, as the LTL teachers did not talk explicitly about their LS team members' perspectives, in contrast with HTL teachers.

In practice, facilitators could prioritize helping teachers understand what LS is by providing enough time for them to notice the value of learning through interaction with colleagues. The facilitator could continuously make explicit *how* interaction benefits their learning. Participants need to understand that they can build on one another's knowledge and that they can challenge their own assumptions on student learning and teaching by making use of one another's perspectives and gained expertise.

Future research could explore why some LS participants show more or less interest in learning from their colleagues, investigating such things as the establishment of a safe learning climate and teachers' beliefs about (collaborative) professional development for teachers and beliefs about education in general (De Vries et al., 2013). The findings also raise the question whether a causal relationship exists between understanding of LS and attitude toward LS; this relationship could be a worthwhile starting point for future research.

We acknowledge three limitations of this study. The first pertains to the limited insights some variables provided into our understanding of teacher learning due to lack of variety between the given answers, too much variety between the answers, or participants' difficulty in reflecting on the questioned variable. On the scales for focus on student learning and preconditions of communication, the HTL and LTL participants answered similarly, and the interview data did not provide additional information. This lack of variety and additional interview data made it impossible to determine whether the different scales influenced teacher learning, although they both seem important for teacher learning. Future research could work to gain more understanding about these variables and ask more specific questions about these topics during the interviews.

By contrast, participants' responses about understanding of LS had too much variation, making the results too arbitrary on a detailed level, though they are insightful on an overarching level. We find that LS can be explained from multiple angles and layers, as reflected in our interview data. Therefore, for future research, we recommend thoroughly thinking through how to narrow down interview questions on the understanding of LS when conducting similar studies.

Regarding participants' difficulty in reflecting on the questioned variable, we found that they shared relatively few insights into the characteristics of the conversations among LS team members. Thus, the question arises whether questions on characteristics of the conversations require too specific metacommunicative knowledge and jargon with which participants are unfamiliar. Conversational characteristics do seem to be of great importance for teacher learning during LS (e.g., Amador & Weiland, 2015; Dudley, 2013; Vrikki et al., 2017). We recommend that future research evaluate the use of exploratory talk by analyzing recorded LS conversations instead of asking for participants' perspectives.

A second limitation of our study is that we assumed that LS participants conduct the LS actions collaboratively. However, during the interviews two teachers shared that the research lesson was developed by one teacher rather than the whole LS team. Whether one teacher or the entire team develops the research lesson could be of significance for experienced teacher learning. In future studies, we advise considering a more open view on distribution of LS tasks among the LS participants and incorporating questions on how the tasks are divided not only between LS participants and the facilitator but also among participants.

Third, we address limitations due to the small sample size, for both the quantitative and qualitative parts of this study. For the quantitative analyses, we used an independent *t*-test with bootstrapping, which is considered an effective way to analyze small data samples (Dwivedi et al., 2017). Despite the low power of data from small sample sizes to show significant differences between groups, we observed that LTL teachers scored consequently lower on every scale in the questionnaire. This finding strengthens our conclusion that the theorized variables are indeed of relevance in relation to teacher learning. However, overall the possibilities for statistical approaches were limited. Studying correlations and performing other analyses (e.g., multivariate analyses, cluster analysis) on the questionnaire data could provide insight into the distinguishing influences on teacher learning during LS. An increased sample size would also enhance variety between the participants, which would allow researchers to examine differences between and within groups to gain more insights into causality and what is needed given specific contexts.

Finally in relation to the small sample size, we conducted the qualitative analysis with only five HTL and two LTL teachers from the same school context. Future research could include more HTL and LTL teachers from the same as well as different school contexts to gain more insights into which influences are generic and which are context specific.

In general, this study provides an overview of variables that are relevant for researchers to focus on in researching how LS works, as well as focal points for school and LS facilitators when implementing or evaluating LS and suggestions for how to strengthen LS cycles. This study specifically illustrates the importance of enhancing teacher learning through LS by focusing on the alignment among what teachers want to learn from LS, their research question, and the research lesson to ensure that LS conducted is relevant to them. This study also illustrates the importance of stimulating the use of external expertise, supporting teachers in forming an understanding of how LS contributes to their learning, helping teachers understand what LS is and is not, and supporting teachers in developing a positive attitude toward leaning from colleagues through interaction.

Funding

This study was funded by the Nationaal Regieorgaan Onderwijsonderzoek [National Governing Body Educational Research] [405-17-810, 2017–2022].

Acknowledgment

We would like to thank the participants for sharing their time and experiences with us and the LS coordinator at the school for helping such a great deal in organizing the data collection. We would also like to thank Sofie van Harskamp for her earlier active involvement in the development of the preliminary version of the used questionnaire.

Appendix

Example of a vignette

Vignette 5 – Post-lesson discussion.

The LS team has scheduled 30 min for the post-lesson discussion. They have observed a research lesson with the following learning goal: "How can we teach students to substantiate their opinion to others?" The teachers discuss whether the learning goal for their observed case student was achieved. They all indicate that the students achieved the learning goal and use their observations to collaborate this, such as "they were really constructive towards each other, really nice to watch. They gave each other compliments, tips and tricks. My case student participated well beyond expectations; I expected that she would not like it.'"

Note. The vignette contains some common misinterpretations

about (how to conduct) aspects of LS to prompt interviewees to explain how they understand LS.

References

Amador, J., & Weiland, I. (2015). What preservice teachers and knowledgeable others professionally notice during lesson study. *The Teacher Educator, 50*, 109–126. https://doi.org/10.1080/08878730.2015.1009221

Arminen, I. (2005). *Institutional interaction: Studies of talk at*. Ashgate Pub.

Assen, J. H. E., & Otting, H. (2022). Teachers' collective learning: To what extent do facilitators stimulate the use of social context, theory, and practice as sources for learning? *Teaching and Teacher Education, 114*, Article 103702. https://doi.org/10.1016/j.tate.2022.103702

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

Choy, B. H., & Lee, C. K. E. (2021). Going deeper into lesson study through kyouzai kenkyuu. In A. MurataA, & C. K. E. Lee (Eds.), *Stepping up lesson study: An educator's guide to deeper learning* (pp. 39–51). Routledge. https://doi.org/10.4324/9781003002536.

Choy, B. H., Thomas, M. O. J., & Yoon, C. (2007). The FOCUS framework: Characterizing productive noticing during lesson planning, delivery and review. In E. O. Schack, M. H. Fisher, & J. A. Wilhelm (Eds.), *Teacher noticing: Bridging and broadening perspectives, contexts, and frameworks* (pp. 445–466). Springer. https://doi.org/10.1007/978-3-319-46753-5_26.

De Vries, S., Jansen, E. P., & van de Grift, W. J. (2013). Profiling teachers' continuing professional development and the relation with their beliefs about learning and teaching. *Teaching and Teacher Education, 33*, 78–89.

De Vries, S., & Roorda, G. (2019). Het leren van docenten in een lesson study team: Een casestudie [the learning of teacher in a lesson study team: A case study]. *Pedagogische Studies, 96*(6), 401–422.

De Vries, S., Roorda, G., & Van Veen, K. (2017). *Lesson Study: Effectief en Bruikbaar in het Nederlandse Onderwijs? [Lesson Study: effective and useable in Dutch education?* University of Groningen.

De Vries, S., & Uffen, I. (2021). Facilitating a lesson study team to adopt an inquiry stance. In A. Murata A, & C. K. E. Lee (Eds.), *Stepping up lesson study: An educator's guide to deeper learning* (pp. 94–105). Routledge.

De Vries, S., Verhoef, N., & Goei, S. L. (2016). *Lesson study: Een praktische gids voor het onderwijs [lesson study: A practical guide for education*. Garant.

Desforges, C. (1995). How does experience affect theoretical knowledge for teaching? *Learning and Instruction, 5*, 385–400. https://doi.org/10.1016/0959-4752(95)00024-0

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

Devers, K. J. (1999). How will we know "good" qualitative research when we see it? Beginning the dialogue in health services research. *Health Services Research, 34*(5), 1153–1188.

Dudley, P. (2011). *Lesson study: A handbook*. Retrieved from http://lessonstudy.co.uk/wp-content/uploads/2012/03/new-handbook-revisedMay14.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. (2015). How lesson study works and why it creates excellent learning and teaching. In P. Dudley (Ed.), *Lesson study: Professional learning for our time* (pp. 1–24). Routledge. https://doi.org/10.4324/9780203795538.

Dudley, P., Xu, H., Vermunt, J., & 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: Research, Development and Policy, 54*(2), 202–217. https://doi.org/10.1111/ejed.12337

Dwivedi, A. K., Mallawaarachchi, I., & Alvarado, L. A. (2017). Analysis of small sample size studies using nonparametric bootstrap test with pooled resampling method. *Statistics in Medicine, 36*(14), 2187–2205. https://doi.org/10.1002/sim.7263

Ehren, M. C. M. (2006). *Toezicht en schoolverbetering [Supervision and school improvement*. [Dissertation, University of Twente.

Engeström, Y. (1987). *Learning by expanding. An activity-theoretical approach to developmental research*. Orienta-Konsultit.

Engeström, Y., & Sannino, A. (2010). Studies of expansive learning: Foundations, findings and future challenges. *Educational Research Review, 5*(1), 1–24. https://doi.org/10.1016/j.edurev.2009.12.002

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.

Fujii, T. (2016). Designing and adapting tasks in lesson planning: A critical process of lesson study. *ZDM Mathematics Education, 48*(4), 411–423. https://doi:10.1007/s11858-016-0770-3.

Hargreaves, A. (2001). The emotional geographies of teachers' relations with colleagues. *International Journal of Educational Research, 35*(5), 503–527. https://doi.org/10.1016/S0883-0355(02)00006-X

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

Koffeman, A., & Snoek, M. (2018). Identifying context factors as a source for teacher professional learning. *Professional Development in Education, 45*(3), 456–471. https://doi.org/10.1080/19415257.2018.1557239

Lee, C. K. E. (2015). Examining education rounds through the lens of lesson study. *International Journal of Educational Research, 73*, 100–106. https://doi.org/10.1016/j.ijer.2015.07.001

Lee Bae, 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

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

Lewis, C. (2005). How do teachers learn during lesson study? In P. Wang-Iverson, & M. Yoshida (Eds.), *Research for better schoolsBuilding our understanding of lesson study* (pp. 77–84).

Lewis, J. M. (2016). Learning to lead, leading to learn: How facilitators learn to lead lesson study. *ZDM Mathematics Education, 48*(4), 527–540. https://doi.org/10.1007/s11858-015-0753-9

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. Takahasi, & J. P. da Ponte (Eds.), *Theory and practice of lesson study in mathematics* (pp. 13–37). Springer. https://doi.org/10.1007/978-3-030-04031-4_2.

Lewis, C., & Perry, 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. https://doi.org/10.1007/978-3-319-07716-1_7.

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

Lincoln, Y. S., Lynham, S. A., & Guba, E. G. (2005). Paradigmatic controversies, contradictions, and emerging confluences revisited. In N. K. Denzin, & Y. S. Lincoln (Eds.), *The SAGE handbook of qualitative research* (4th ed., pp. 97–128). SAGE publications.

Mercer, N. (1995). *The guided construction of knowledge. Talk amongst teachers and learners*. Multilingual Matters.

Miedema, W. G., & Stam, M. (2008). *Leren van innoveren: Wat en hoe leren docenten van het innoveren van het eigen onderwijs? [Learning from innovation: What and how do teachers learn from innovating their own education?* Van Gorcum.

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

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

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

Roth, W. (2007). Emotion at work: A contribution to third-generation cultural-historical activity theory. *Mind, Culture and Activity, 14*(1–2), 40–63. https://doi.org/10.1080/10749030701307705

Roth, W., Lee, Y., & Hsu, P. (2009). A tool for changing the world: Possibilities of cultural-historical activity theory to reinvigorate science education. *Studies in Science Education, 45*(2), 131–167. https://doi:10.1080/03057260903142269.

Salas, E., Sims, D. E., & Burke, C. S. (2005). Is there a "big five" in teamwork. *Small Group Research, 36*, 555–599. https://doi.org/10.1177/1046496405277134

Sannino, A. (2011). Activity theory as an activist and interventionist theory. *Theory & Psychology, 21*(5), 571–597. https://doi.org/10.1177/0959354311417485

Schipper, T. M. (2019). Teacher professional learning through lesson study: An examination of lesson study in relation to adaptive teaching competence, teacher self-efficacy, and the school context. *Rijksuniversiteit Groningen*. https://doi.org/10.33612/diss.98636764

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:10.1108/IJLLS-10-2016-0041.

Stepanek, J., Appel, G., Leong, M., Mangan, M. T., & Mitchell, M. (2007). *Leading lesson study. A practical guide for teachers and educators*. Corwin Press.

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

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*, 2–7.

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

Takahashi, A., & Yoshida, M. (2004). Ideas for establishing lesson-study communities. *Teaching Children Mathematics, 10*, 436–443. https://doi.org/10.5951/tcm.10.9.0436

Uffen, I., De Vries, S., Van Harskamp, S., & Goei, S. L. (2019). *Facilitating teachers to use different perspectives during lesson study conversations: Predicting the effect on teacher learning.* April 5-9. Toronto, Canada: AERA annual meeting [Paper presentation].

Van Driel, J. H., Meirink, J., Van Veen, K., & Zwart, R. (2012). Current trends and missing links in studies on teacher professional development in science education: A review of design features and quality of research. *Studies in Science Education, 48*(2), 129–160. https://doi.org/10.1080/03057267.2012.738020

Van Eekelen, I. M., Vermunt, J. D., & Boshuizen, H. P. A. (2006). Exploring teachers' will to learn. *Teaching and Teacher Education, 22*(4), 408–423. https://doi.org/10.1016/j.tate.2005.12.001

Van Harskamp, S. (2018). *A teachers' Perspective on the Role of the Facilitator in lesson study [Master's thesis.* University of Groningen.

Van Oers, B. (2009). Narrativiteit in leerprocessen. Ontwikkelingen binnen de cultuurhistorische benadering van het leren [Narrativity in learning processes. Developments within the cultural-historical approach to learning]. *Pedagogische Studies, 86*(2), 147–156.

Van Veen, K., Zwart, R. C., Meirink, J. A., & Verloop, N. (2010). *Professionele ontwikkeling van leraren: een reviewstudie naar effectieve kenmerken van professionaliseringsinterventies van leraren [Professional development of teachers: a review study into effective characteristics of teacher professionalization interventions].* ICLON/Expertisecentrum Leren van Docenten.

Van den Boom - Muilenburg, S. N., de Vries, S., van Veen, K., Poortman, C. L., & Schildkamp, K. (2021). Understanding sustainable professional learning communities by considering school leaders' interpretations and educational beliefs. *International Journal of Leadership in Education*. https://doi.org/10.1080/13603124.2021.1937705

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

Wei, G. (2019). How could cultural-historical activity theory inspire lesson study? In R. Huang, A. Takahashi, & J. da Ponte (Eds.), *Theory and practice of lesson study in mathematics, advances in mathematics education* (pp. 39–58). Springer. https://doi.org/10.1007/978-3-030-04031-4_3.

Wolthuis, F., van Veen, K., de Vries, S., & Hubers, M. D. (2020). Between lethal and local adaptation: Lesson study as an organizational routine. *International Journal of Educational Research, 100*. https://doi.org/10.1016/j.ijer.2020.101534