Teacher learning in the context of Lesson Study: A video-based analysis of teacher discussions

Maria Vrikki[1], Paul Warwick[1], Jan D. Vermunt[1], Neil Mercer[1], & Nicolette Van Halem[2]

[1]University of Cambridge, Faculty of Education, 184 Hills Road, Cambridge, CB2 8PQ, United Kingdom
[2]University of Utrecht, Faculty of Social and Behavioural Sciences, PO Box 80125, 3508 TC Utrecht, The Netherlands

***Corresponding author: mv387@cam.ac.uk; Telephone +44 1233 767569**

Highlights:

- The development of a reliable coding scheme for teacher learning is reported.
- Learning in a group has an impact on teachers' individual learning processes.
- Dialogic moves in Lesson Study discussions are a mechanism for learning.
- Descriptive and interpretative learning processes are observed in Lesson Study.

Abstract

This paper contributes to our understanding of teacher learning in the context of Lesson Study (LS), a model of professional development that involves collaborative lesson planning and evaluation. Video-recorded LS discussions of mathematics teachers based in London were analysed for this purpose. Two inter-related studies are presented: the first involved the construction of a reliable coding protocol for video analysis; the second used this protocol for coding 120 fragments of discussions amongst 91 teachers. Findings are discussed with reference to tests of reliability and results of multilevel analysis, which reveal differential effects of particular forms of interactions on learning processes.

Keywords: Teacher learning; Professional development; Dialogue; Lesson Study.

1. Introduction

Despite some relatively recent work in the discipline of mathematics (see Section 1.1), research on in-service teacher *learning* is in its infancy. This is in contrast to studies on the relative effectiveness of teacher professional development (PD) (e.g. Guskey, 2002), of which there are many. Definitions of learning, and internal mechanisms for learning, have been debated for decades. For the purposes of this paper, learning is seen as a change or development in knowledge, resources or understanding that have the potential to lead to professional behavioural change. In considering teacher learning we adopt a sociocultural perspective, seeing the mechanism as "the dynamic interdependence of social and individual processes" (John-Steiner & Mahn, 1996, 192), with language as the central cultural tool in facilitating this interdependence and expressing developed understanding (the connection with language is explored in Section 1.3). This implies that learning takes place as the result of interactions between individuals, or between individuals and cultural tools, with knowledge within groups often being co-constructed as a result of spoken interactions.

When considering learning in groups, many researchers see the environment and social structures as key to the cognitive activities associated with collaboration (Dillenbourg, 1999). Following from this, a prominent finding of existing research into teacher learning is that professional communities are effective 'agents' for enhancing professional learning and sustained professional development (Stoll, Bolam, McMahon, Wallace, & Thomas, 2006; Webster-Wright, 2009). Indeed, schools with strong teacher communities seem to have higher student achievement (Horn & Kane, 2015; Bryk, Sebring, Allensworth, Luppescu, & Easton, 2010).

However, Webster-Wright (2009) suggests that little is understood about the effective mechanisms of learning in such professional communities. 'Black box' models of teacher learning thus suggest teacher communities have an effect on teacher learning, but not how that effect is brought about.

It is the purpose of this paper to explore the processes of teacher learning within these communities. We focus on the role of dialogue and draw on the research into the effective use of talk in group contexts. We are interested in how talk is being used to foster learning in professional groups and we pursue this interest through studying teachers' discussions that occur in the context of Lesson Study (LS), a model of PD now employed in many countries around the world (Dudley, 2013; Lee, 2011). In so doing we demonstrate relationships between three fields of study – teacher learning, dialogue in education and professional settings, and Lesson Study (Figure 1). In this, LS is distinct from the other two fields in one important respect; it is a specific methodology intended to improve student outcomes, rather than an argued theoretical domain, as teacher learning and dialogue might be viewed (Niss, Peng Yee and Kilpatrick, 2015).

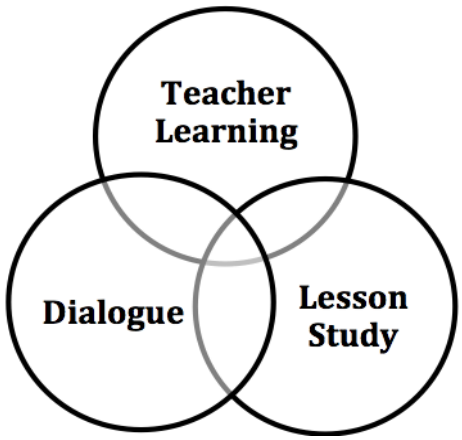

Figure 1. Combination of three theoretical perspectives

1.1. Teacher learning and teacher professional development

As we have stated, literature on teacher learning is largely situated in two separate contexts: that of pre-service teachers and that of in-service teachers, with an imbalance between the level of work conducted in each context clearly evident. While the field of pre-service teacher learning has received much attention, the field of in-service teacher learning - the focus of the present paper - is emergent. In the field of pre-service teacher learning, substantial work has been done on the learning of mathematics teachers and in particular on the types of knowledge that they should possess. Shulman (1987) identified seven types of teacher knowledge, placing particular emphasis on three types with content-specific dimensions: content knowledge, curriculum knowledge and pedagogical content knowledge. Later work in mathematics education built on Shulman's work by identifying or extending types of knowledge. This work has included Ball et al's *Mathematics Knowledge for Teaching* (Ball, Thames & Phelps, 2008), and Rowland et al's *Knowledge Quartet* of foundation, transition, connection and contingency (Rowland, Huckstep and Thwaites, 2005). One consensus is that "knowledge of *mathematics, didactics* and *pedagogy*" (Jaworski & Huang, 2014, 175) are necessary for mathematics teachers.

In addition to such work, substantial work has been concerned with the ways in which teachers can develop their professional knowledge. This is particularly important in the context of the work reported here. Our study is situated in the context of the introduction of the new National Curriculum for Mathematics in England (September 2013). This reform stressed the importance of developing skills such as mathematical reasoning and proof, and having mathematical fluency. Here, we focus on teacher LS discussions about how to develop the teaching of these skills, rather than on the varied mathematical content knowledge that was the context for such skills development.

Thus, through the detailed case observations required in LS (Section 1.2), the central work of our teacher groups was to develop Knowledge of the interaction between Content (including mathematical skills) and Students (KCS - Ball et al., 2008). The subject areas within mathematics that the students' work focused on were many and varied.

In considering the ways in which teachers can develop their professional knowledge, reflective practice is seen as key. As many authors argue, developing practitioners who are constantly *reflecting on-action* and *in-action* (Schön, 1987) is fundamental (Jaworski & Huang, 2014), as it allows teachers to develop a critical lens (Cochran-Smith, 2003) in their work. Nowadays, *communities of practice* (Lave & Wenger, 1991) or "*inquiry communit[ies]"* (Jaworski, 2008, p. 312) are a widely used means that enable teachers to learn in and from practice. They are made up of colleagues who share an understanding of school culture and have common interpretations of their intentions (Matos, Powell & Sztajn, 2009). Participating in such communities allows teachers to *co-learn* (Jaworski, 2001, 2003) by developing situated learning through critical evaluation of their practice. Reflection thus becomes a social endeavour, rather than an individual, internal process. While the substantial work in this field offers insights on what and how teachers develop professional knowledge, work on the specific thinking processes that enable this learning is still limited. As Matos et al. (2009) state, 'research on learning shows that we need languages to describe in analytical terms the process of coming to know' (171).

Considering specifically in-service teacher learning, until recently this has been measured mainly by the 'effectiveness' of teacher professional development programmes. Several review studies demonstrate the ways in which these have been

measured (e.g. Borko, Jacobs & Koellner, 2010; Postholm, 2012). In a recent review, Van Driel, Meirink, Van Veen and Zwart (2012) used Desimone's (2009) analytic framework (Figure 2) in order to examine how the effectiveness of PD programmes for science teaching had been measured in previous research.

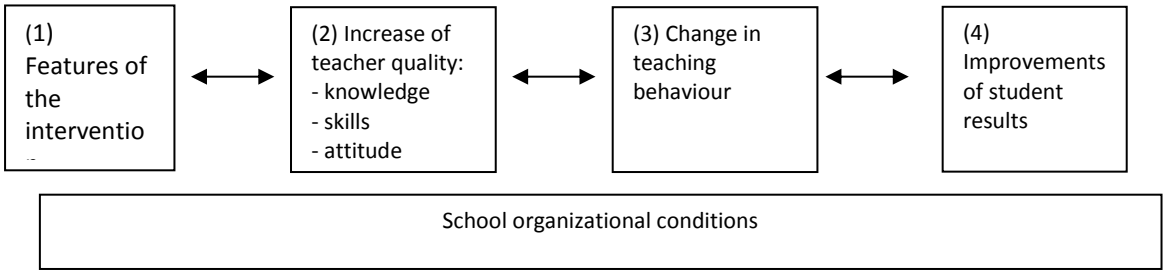

Figure 2. Analytical framework for studying the effectiveness of teacher professional development (Van Veen, Zwart & Meirink, 2012, based on Desimone, 2009).

From the 44 studies that met their inclusion criteria, Van Driel et al. (2012) found that four studies (9%) measured the effectiveness of the PD programmes based on the relationship between the intervention and changes in teachers' cognitions, i.e. knowledge (1+2 in the model, Figure 2); three studies (7%) looked at the relationship between the intervention and changes in teachers' classroom behaviour (1+3 in the model); half examined the relationship between the intervention and changes in both teacher cognitions and classroom behaviour (1+2+3); and fifteen studies (34%) examined all four aspects of the model. However, they differed in the way they measured student outcomes with six studies (14% of the total of 44 studies) using achievement tests and nine using teachers' views on student progress.

The analytical model is therefore an intervention-outcome model. It conceptualizes teacher learning as features of the intervention leading to teacher learning outcomes (i.e. knowledge, skills, attitudes), which can then lead to changes in teacher behaviour

in the classroom. However, it does not address the processes of teacher learning. Such PD programmes are thus 'black box models' because they do not make the processes between stimulus (intervention) and response (learning outcomes) explicit (Author, 2013).

In contrast, the field of student learning has a large body of research focusing on learning processes. In particular, Marton and Säljö (1984) identified two distinct learning approaches: a deep approach and a surface approach. A deep approach is characterised by the student's intention to understand; activities might include looking for relations between ideas, structuring the subject matter and critical engagement with literature. A surface approach, on the other hand, is characterised by the student's intention to remember the learning material, rehearsing ideas in order to be able to reproduce them. Other researchers have built on the findings by elaborating and extending their conceptualisations (see Author, 2004; Richardson, 2000; Entwistle & McCune, 2004; Lonka, Olkinuora & Mäkinen, 2004; Pintrich, 2004). Author (2004) use the term 'learning pattern' as an encompassing concept in which the cognitive processing of subject matter, the metacognitive regulation of learning, conceptions of learning and learning orientations are united. In studies with university students they found four such patterns of learning: undirected, reproduction directed, meaning directed and application directed (see also, for example, Authors, 2014).

One study concerned with both *what* and *how* teachers learn was that of Authors (2010) that took place in a context of educational innovation focused on active and self-regulated learning. Ninety-four participating teachers sent digital logs to the researchers six times in the course of one year describing a recent learning experience. Using a total of 735 learning experiences, Authors' (2010) content

analysis revealed six categories of learning activities, with the four most frequent ones being: (1) experimenting (32%); (2) considering one's own practice (34%); (3) Getting ideas from others (15%); and (4) experiencing 'friction' (15%).

In the research reported here, we have taken the investigation of what is inside the 'black box' of teacher learning a step further by examining how that learning is related to teachers' engagement in LS discussions.

1.2. Lesson Study

Lesson Study (LS) *seems* to provide an ideal platform for teacher learning to take place. It is a teacher-led PD model, which involves teachers' collaborative planning and evaluation of research lessons that focus particularly on student learning. Through the LS 'platform', the learning activities found in the Authors (2010) study should be fostered, as it allows teachers to experiment, reflect on their practice by analysing student behaviour, 'interthink' with other LS members and reveal discrepancies between their expectations and observations.

Known as *jugyo kenkyu* (Lewis & Tsuchida, 1997), LS originated in Japan in the 1870s after being introduced to the country by Western educators. After years of implementation, several curriculum and teaching theories (Isoda, 2015) have emerged through LS in Japan, and the methodology is considered to represent "the culmination of innovations and improvements" (Shimizu & Chino, 2015, p. 123). LS comprises of three main steps: 1) *Plan*, involving teachers preparing and studying lessons; 2) *Do*, involving the teaching and observations of lessons; and 3) *See*, involving post-open class discussion and reflection among teachers and sometimes others (Isoda, 2015).

Although variations of this process are found internationally, these three steps remain the core of all adapted models.

LS became popular worldwide after the publication of Stigler and Hielbert's (1999) *The Teaching Gap*, which demonstrated that Japanese students' achievement in mathematics surpassed that of students in other countries, citing LS as a possible explanation for this success. The highly adaptable core model enabled its adoption by teachers of different subjects across a range of different contexts, including mathematics (e.g. Murata, Bofferding, Pothen, & Taylor, 2012; Meyer & Wilkerson, 2011; Cajkler, Wood, Norton & Pedder , 2013); English (e.g. Lee, 2008); geography (e.g. Cajkler et al., 2013); and science (e.g. Chong & Kong, 2012). Such studies reveal that LS incorporates many of the elements identified by Borko et al. (2010) as characteristics of the most effective PD:

1) content is situated and addresses problems of practice;
2) content is focused on students' learning;
3) preferred instructional practices are modelled in the PD;
4) active learning and inquiry by teachers are key elements.

Research to date has yielded promising findings in terms of teacher learning outcomes. Lewis, Fischman, Riggs and Wasserman (2013) found that the elementary and middle schools mathematics teachers, who participated in LS in the US, expanded their mathematical knowledge, became more curious about analyzing students' thinking and made more use of multiple representations for solving mathematics problems. Similarly, in a qualitative study of a LS group of two mathematics coaches and one classroom teacher, Knapp, Bomer and Moore (2011) found that their involvement led to knowledge gains about ways of teaching, students' learning and

specialized subject content. A case study of mathematics teachers (Kamina & Tinto, 2011) revealed that LS participants developed their understanding of subject content and their pedagogic skills, and encouraged more positive beliefs about students' learning. Another qualitative study conducted with ten teachers in a Singapore high school concluded that LS participation led to an "increase in teachers' knowledge, perspective and insights about instruction and subject content" (Lawrence & Chong, 2010, p. 567). Finally, Meyer and Wilkerson (2011) worked with 24 middle school mathematics teachers in LS groups. Transcripts of teachers' discussions revealed an increase of the participants' knowledge of teaching mathematics.

Despite these promising findings of "the potential of [LS] to promote teacher change" (Tepylo & Moss, 2011, p. 59), the evidence has come only from small-scale, qualitative investigations, which do not examine the mechanisms that lead to improved teaching – the *how* or *why* LS works (Corcoran, 2011). Large-scale studies that have this focus are needed and the research reported here was designed to address this need. In order to do so, we focus on identifying productive discussions in teachers' LS meetings, which we suggest are the driving mechanism of LS. Whilst the LS process as a whole can be a source of new knowledge for teachers (i.e. they can learn from research lessons and pupil interviews), we focused our attention on how collaborative spoken dialogue creates opportunities for them to learn.

1.3. Dialogue

Rooted in a sociocultural framework of understanding learning, a focus on dialogue reflects our wish to understand the processes by which teachers exchange information and build professional knowledge in collaborative groups; in this conception,

dialogue represents the observable manifestation of thinking. Research into collaborative working, across a range of professional contexts, suggests that "groups seem to achieve some of the best, and some of the worst, outcomes" (Authors, 2013, p. 24). Analysing the quality of communication within a group can provide an explanation of why some groups achieve productive outcomes whilst others flounder. The increased resource of knowledge and expertise available when people work together in groups *can* lead to more productive and creative solutions to problems. The existence of a common conception of the shared goal, the drawing together of knowledge relevant to that goal, and the availability of shared concepts vocabulary *should* help professional working groups make progress. Housley (2000) argues that "knowledge as a situated phenomenon is one which those interested in promoting dialogue and the exchange of information between team members or individuals within groups should consider" (p. 104).

However, research has shown that group discussion is far from reliable in generating positive outcomes. The implicit social norms of discussion in some professional groups may be used to strictly control how knowledge exchange, knowledge building, and particularly dissent around ideas, are handled (Author, 2000; Authors, 2013). Their purpose can be to stifle dissent and enable the group to work quickly towards uncontested consensus, rather than to find the best possible solution to a problem. Labelled 'groupthink' (Janis, 1982; Hart, 1994; Esser, 1998), this phenomenon has been blamed for some catastrophic political decisions. So group discussions that quickly narrow the conceptual space around what is being discussed may be more efficient, but can be ineffective in terms of achieving positive outcomes (Middup, Coughlan, & Johnson, 2010).

In contrast, professional groups that regularly produce good solutions to problems seem to do so because of "ground rules' that generate positive conditions for interaction and knowledge building; they seem to act in what 'may be imagined as a socially distributed cognitive process" (Måseide, 2003, p.369). The talk and interaction in such professional groups mirror that found in the research on effective group interaction amongst school students (Authors, 2007). Nemeth, Rogers, and Brown (2001), for example, found that a wider range of relevant perspectives were taken into account where a group member or members expressed doubts about proposals or ideas being expressed by the majority. It seems to be both the features of group dialogue, and the supportive environment that enables such dialogue, that are of importance.

In everyday use, dialogue means conversation, talking together. But within a sociocultural framework of understanding (Vygotsky, 1962; Bakhtin, 1981), dialogue refers to "any kind of human sense-making, semiotic practice, interaction, thinking and communication, as long as these phenomena are 'dialogically' (or 'dialogistically') understood" (Linell, 2009, p. 990). Dialogue can be thought of as a very specific use of a central cultural tool, enabling people to come to an understanding of one another's knowledge and perspectives through 'interthinking' (Author, 2000). In Vygotskian terms, children learn ways of thinking through 'intermental' functioning (i.e. interaction with other people) and this shapes their 'intramental' functioning. However, there is no reason to believe that as we become adults, we cease any 'intermental' functioning. Indeed, interaction for intellectual purposes, in the pursuit of joint goals, is a common and essential feature of adult human life and one that is particularly prevalent in well-run professional groups.

From previous analyses of classroom-based interactions we draw some key understandings to inform our analysis of professional dialogue. A dialogic pedagogy involves teachers and learners in: commenting and cumulatively building on each other's ideas, posing questions and constructing interpretations together (Alexander, 2011); 'seeing' things from someone else's perspective; 'chaining' questions and answers; and constructing and critiquing shared knowledge (Barnes & Todd, 1977; Author, 2000). Research (Authors, 2007; Littleton & Howe, 2010; Wegerif, 2007) highlights a need for open questions in dialogic exchanges, which stimulate extended responses and reasoning. There is now robust evidence (Authors, 2004a; Rojas-Drummond, Littleton, Hernández, & Zúniga, 2010) showing a direct effect on individual and collective outcomes (Alexander, 2012). In a systematic review covering 40 years of research, Howe and Abedin (2013) found positive associations between student learning and extended and cumulative responses in group dialogue, highlighting the importance of expressing competing viewpoints and working towards agreement. In the light of such work, professional initiatives for the promotion of classroom dialogue have been developed (Authors, 2003; Alexander, 2011; Anderson, Chinn, Waggoner, & Nguyen, 1997; Myhill, Jones, & Hopper, 2006).

It seems then that dialogue provides a mechanism for critical thinking and reasoning, a framework for productive collaboration and communication, and a means by which desirable generic competences can be achieved. This study set out to establish the pertinence of such ideas to the professional discussions carried out by teachers in the context of LS. In devising an analysis protocol for examining such discussions, we were concerned with identifying the dialogic moves that most reliably seemed to lead to teacher learning.

1.4. The 'Teacher Learning and Lesson Study in Mathematics Higher Order Teaching and Learning' project

We report here on a large-scale two-year project, based upon collaboration between the School Improvement Services of the Borough of Camden in London, UK, and the University of Cambridge. Designed to take account of the introduction of the new National Curriculum for Mathematics in England (September 2013), the project had two parts: 1) professional development which aimed to introduce LS into the practice of communities of mathematics teachers, increase their understanding of the content of the new curriculum, and encourage new relevant pedagogic practices (August 2013-August 2015); 2) research which aimed to develop an understanding of teachers' learning through an analysis of teachers' discussions (February 2014-December 2015). For the latter part, teachers video-recorded their LS meetings on school equipment.

During the two school years of the development project, six LS phases were planned (three each school year) with 59 primary, secondary and special schools participating in as many LS phases as they could. Some schools entered in the first year (Cohort 1) and continued on to the second year; others joined in the second year (Cohort 2). In each school, a group of teachers of Mathematics, usually three, was formed. The Lesson Studies in this project focused on Years 5 and 6 (students aged 10-11, the final years of primary education) and Years 7 and 8 (students aged of 12-13, the first years of secondary education). In some cases the teacher LS groups stayed intact across different LS phases; more often, new teachers replaced others in an attempt to scale up LS practice within schools.

In each phase, Dudley's (2013) three-cycle LS model was implemented (see Figure 3). In groups of three or four, teachers held a planning meeting for the first research lesson (RL1). Decisions included which class to teach, who from the group would teach it and the focus of teaching. As with learning study, lesson study is intended to highlight "the learning problems of the students" (Runesson, 2008, p. 153). Three case pupils were selected as representative of different student groups (Lee, 2011). Specific activities were planned and success criteria were set for each case pupil. The RL was taught by one teacher, the others observing case pupils. After the lesson, the case pupils were interviewed about their views on the lesson. The teachers then held a post-lesson discussion meeting, reflecting on the lesson and making decisions leading to the planning of the second research lesson (RL2). The same steps were followed for the second and third cycle. At the end of the three LS cycles, teachers were encouraged to share any new insights about teaching and learning in some way (e.g. public research lessons, case study report) with interested audiences (within and beyond their schools).

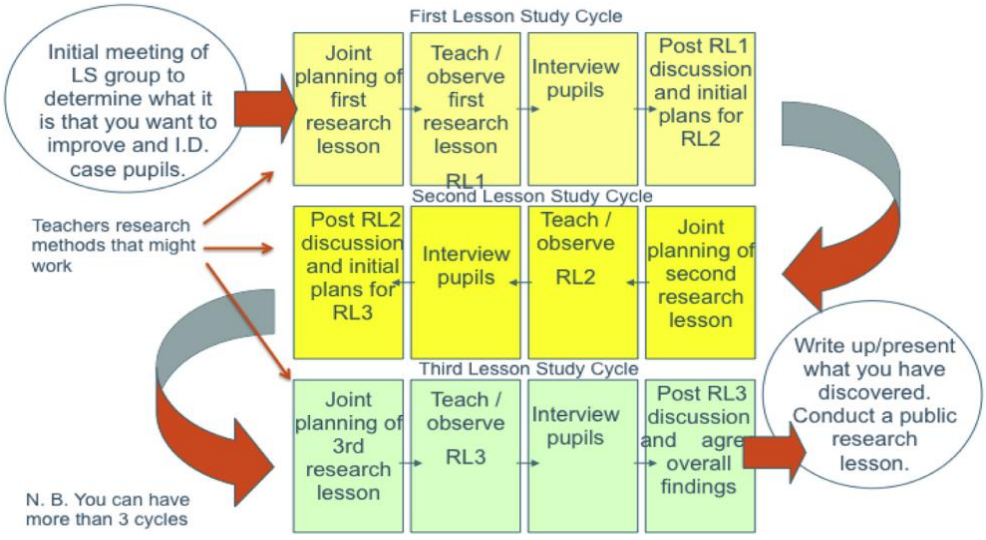

Figure 3. Lesson Study model as used in the project (after Dudley, 2013)

Planning and feedback conferences were organized for the participating teachers by Camden's School Improvement Services at the beginning and end of each LS phase. At these events, teachers shared latest experiences and findings with colleagues from other schools. A mathematics specialist helped teachers familiarize themselves with the new curriculum and discover ways of implementing it in their daily practice. Again, it is worth re-stating that the curriculum did not expect teachers to develop their mathematical content knowledge; rather, it required them to adapt their teaching practices to achieve the goals of developing their students' mathematical reasoning and their ability to use mathematical proof and fluency. In addition to these conferences, a LS workbook, prepared by the School Improvement Services, guided teachers' discussions during the LS meetings. It included space for lesson plans, descriptions of expectations for case pupils, suggested questions for pupil interviews, and questions guiding reflective discussions.

The professional development element of the project aimed to facilitate teachers' learning. The research element of the project investigated *how* this learning took place and sought to understand the role of spoken dialogue in the collaborative LS learning process. It was clear that the development of a coding protocol would be essential for this purpose and the first part of research activity was devoted to this. The research comprised two related pieces of work, which we have called 'studies'. Both studies are reported here: the 'Instrument Development Study' reports on the development of a reliable, and valid, tool for coding teacher learning in LS meetings; the 'Teacher Learning Study' reports on the use of this reliable tool for coding video data with the aim of understanding learning processes. As the protocol development process became as important as its application, we decided to separate the two studies in order

to highlight this development process. In addition, we envisage the product of the Instrument Development Study to be used for further analyses in other relevant contexts.

The research question that guided the Instrument Development Study was:

> Can a reliable protocol be devised for analysing the relationship between dialogue and teacher learning processes?

In using the protocol, the main research question for the Teacher Learning Study was:

> How is the content and structure of teachers' LS discussions related to the nature and quality of their learning?

Our research data (for both studies) came from the large pool of video recordings of the LS meetings. Table 1 presents the number of LS groups that recorded their meetings for the project, as well as the total number of teachers appearing in the videos of each phase.

Table 1.
Number of LS groups (schools) and teachers participating in each phase of the project

	Number of LS groups that provided videos	Number of participating teachers
Phase 1 (Sep-Dec 2013)	22	61
Phase 2 (Jan-Mar 2014)	22	39
Phase 3 (May-Jul 2014)	5	17
Phase 4 (Sep-Dec 2014)	45	115
Phase 5 (Jan-Mar 2015)	34	85
Phase 6 (May-Jul 2015)	16	48

2. Instrument Development Study

2.1. Methods

2.1.1 Participants

Project data came from four primary schools that joined the project on Year 2. They were randomly selected from the databases of Phases 4 and 5. Participants had no prior LS experience at Phase 4, whereas participants of Phase 5 either had no experience (if they just got involved with the project) or had experience of one LS phase. A total of 13 teachers from the four LS groups were involved (Table 2).

Table 2.
Participants of instrument development study (selected from Project Phases 4 & 5)

	Number of teachers in LS group
School 1 (Primary)	4
School 2 (Primary)	3
School 3 (Primary)	3
School 4 (Primary)	3
Total	13

Ethical approval was sought from the participating schools for their video-recordings to be used for this research.

2.1.2. Materials

Two main types of materials were used in the Instrument Development Study: video material and a pre-existing version of the protocol.

Videos were sampled from Phases 4 and 5 (as suggested in Table 2) because videos from Phases 1 and 2 had been used for previous trial tests of reliability; and Phases 3

and 6 involved cross-phase LS, a ‘special’ context in which teachers from primary and secondary schools formed LS groups.

As for the protocol used, this was a pre-existing version of the protocol, which had been developed in an iterative manner during the earliest phases of the project. This involved using both ‘bottom-up’ and ‘top-down’ approaches for identifying relevant features of observed talk, refining them and grouping them. The team tried to identify all relevant, observable features of talk in the videos that might help to answer the question of how discussion can lead to teachers’ learning. This approach was influenced by the Sociocultural Discourse Analysis methodology (SCDA: Author, 2004), which views language use ‘as a social mode of thinking – a tool for teaching-and-learning, constructing knowledge, creating joint understanding and tackling problems collaboratively’ (Author, 2004: 137). SCDA also stresses the importance of recognising that learning is commonly a temporal, extended process (Author, 2008). Working in tandem on video analysis and protocol development enabled us to gain insights into how thinking might be co-constructed between teachers in the context of LS and how this then formed new knowledge for members of the group.

To explain this analytic process, Table 3 shows the number of codes in each version of the protocol, including the version that was to finally emerge as a result of the work presented here. Our processes of analysis and reliability testing took the number of codes within the protocol from 54 to 7 over the course of the project. Earlier versions of the protocol are not presented or discussed here; the detailed process of the development of the protocol, along with our rationale for making specific decisions, will be presented in a separate paper.

Table 3.
Number of codes on previous versions of the protocol

	Protocol version	Number of codes
Phase 2	Version 1	54
Phase 3	Version 2	25
Phase 4	Version 2 (same version used for initial reliability tests)	/
Phase 5	Version 3	13
Phase 6	Version 4 (final version)	7

Table 4 below presents the final version of the protocol, which consists of 7 codes. These codes represent three main dimensions: *dialogic moves, scope of discussion* and *learning processes.*

Table 4.
Coding Protocol

DIALOGIC MOVES	SCOPE OF DISCUSSION	LEARNING PROCESSES
[DM1] Requesting information, opinion or clarification	[S1] Groups of pupils	[DLP] Descriptive learning processes
[DM2] Building on ideas	[S2] Particular pupils	[ILP] Interpretative learning processes
[DM3] Providing evidence or reasoning		

Dialogic moves represent contributions to dialogue that were most prominent in taking the conversation forward (Authors, 2016). Here we identify three dialogic moves that were significant in our video data:

a) *Requesting information, opinion and clarification:* Refers to when teachers asked clarification questions, invited opinions or reasoning, and negotiated meaning.

b) *Building on ideas:* Refers to when teachers built on ideas and when they came to some agreement after a difference of opinion.
c) *Providing evidence or reasoning:* Refers to when teachers explained their reasoning, or when they illustrated their opinions.

All three dialogue codes represent integral elements of Exploratory Talk, which Authors (2013, p. 16) consider to be "the most educationally effective" type of talk for promoting critical and constructive contributions.

The *scope of the conversation* dimension focused on the content of the conversation. The LS model used here (Dudley, 2013) encouraged teachers to talk about specific students; i.e. the selected, 'case' students.

Finally, the third dimension, *learning processes,* was of particular interest both conceptually and empirically. Two types were eventually identified: *descriptive learning processes (DLP)* and *interpretative learning processes (ILP).* The first refers to co-construction of knowledge at the level of representing a selection of what was known. Examples include describing lessons plans, activities, expectations for pupils and teaching, and observations of pupil learning and teaching. Recent work on teacher noticing (Sherin, Jacobs & Philipp, 2011) indicates teachers have the capacity to attend intentionally to classroom events, and that this capacity develops with experience. The idea of 'professional noticing' links well to Schön's 'reflection in action', which is characterised as "questioning the assumptional structure of knowing-in-action" (Schön, 1987, p.25). Here, the linking sequence might be: strategic noticing preceding immediate reflection and subsequent action, with learning processes discernible when teachers are asked what led them to their actions. DLP in LS teacher

discussions differs from this, in that the noticing takes place to a large extent 'post-event', and often in response to input from other teachers. Here it has more similarities with Schön's 'reflection on action', where items for consideration are selected on the basis of a teacher's 'appreciative systems' – "repertoires of values, knowledge, theories and practice" (Zeichner and Liston, 1996, p.16). Linking back to our initial definition of learning, by selecting and presenting specific evidence teachers are showing how they are marshalling knowledge resources in a strategic manner to move thinking forward in LS discussions. We have used the term DLP to capture such marshalling as evidence of a process, rather than a singular event.

The second type, ILP, refers to contributions that go beyond the level of description and reveal interpretative thinking identified in three forms: (1) evaluating teaching by considering the effectiveness/appropriateness of activities/tasks; (2) evaluating pupil learning or progress, against the success criteria; and (3) diagnosing pupil errors, misconceptions or problems observed in the lessons. In particular, it will become clear that there was a strong relationship between teacher discussions that extended beyond analysis of individual pupils and teachers' ILP. Interestingly, the initial conceptualisations of DLP and ILP were informed by the literature on students' learning, where different learning processes, such as deep and surface learning, have been proposed (Marton & Säljö, 1984). Here, it is clear that surface learning is of a lower 'quality' than deep learning. In this paper, we address the relationship between DLP and ILP in order to see if there is a similar relationship (Section 3.1.3).

Earlier versions of our protocol included codes for teacher learning points (Authors, 2016) based on teachers' explicit reports on learning (e.g. 'Now I understand why he does that') or of expressions of surprise and sudden insights (e.g. 'Wow. This is very

interesting'). However, even in coding the whole data set of videos, it became clear that these moments only summated a very small percentage of the learning that was taking place in LS discussions; and in relation to our randomly selected videos the application of such explicit reports clearly reduced the validity of the codes[1]. Learning manifested as a process, rather than a moment in time, is not manifested in these utterances. For this reason, the two learning processes codes were used as the more suitable outcomes variables, and codes for learning points were dropped due to validity issues.

2.1.3. Procedure

Trial reliability tests began with the 13 codes of Version 3 of the protocol (Table 3) and results were disappointing (k = .0 to .54). Two coders therefore focused on improving reliability, by having joint video analysis sessions and applying four strategies that were aimed towards conceptual and empirical distinctions. These were: (1) clarifying boundaries between codes; (2) merging or deleting infrequent and non-discriminatory codes; (3) illustrating codes using database; and (4) formulating specific observation rules for each code. The application of these strategies reduced the codes to the 7 of the final version (Table 4).

The specific definitions of codes, as well as the observational rules and representative examples, are presented in Appendix 1. The rules for coding were selected through a 'bottom-up' procedure. Coders collected a range of examples from the data and agreed on how those should be coded. Then, coders looked for common features across those examples and selected the features that characterized most examples to

[1] This is not to say of course, that these occasional individual instances are without value or interest, and we hope to explore them in future work.

use as identifiers when creating rules. For instance, coders noticed that most cases of agreed DLP included a quote, while most cases of agreed ILP involved specific words of evaluation or diagnosing (Appendix 1). This bottom-up approach ensured the validity of the coding.

2.1.4. Analysis

For the final reliability test, episodes were coded separately by each coder. Sixteen, two minute, episodes were randomly selected from the videos of the four primary schools (Table 2). In earlier phases of the project, four-minute episodes were used. This was a level of granularity that took into account issues such as pace of activity, nature of coding categories and cognitive load of the coder (Authors, 2016; Angelillo, Rogoff & Chavajay, 2007). As the number of protocol codes was reduced over time, however, focusing on smaller episodes seemed more appropriate in order to attend to variation of observed codes between teachers.

For the analysis, a binary coding system was used for each individual teacher, with the coders deciding whether or not they could see each code in the talk of each teacher in each episode. Due to the complexity of the coding, each episode was watched four times by the coder, once without coding and once focusing on each teacher (or five viewings if four teachers in a group). The selected episodes created a total of 52 data points in our analysis (Table 5).

Table 5.
Data points for reliability analysis

	Number of teachers in LS group	Number of episodes selected	Data points
School 1	4	4	16
School 2	3	4	12
School 3	3	4	12
School 4	3	4	12
Total	13	16	52

At the end of the coding procedure, the codings from the two coders were combined in SPSS and Cohen's kappa (κ) was used to calculate inter-coder agreement for each code.

2.2. Results

The results of the inter-coder agreement test revealed that agreement was high (Table 6) with substantial agreement on four codes and excellent agreement on three codes reaching.

Table 6.
Results of inter-coder agreement

Dimension	Codes	Cohens *K*	Percentage agreement	Interpretation
Dialogic moves	DM1	.84***	92.3	Excellent
	DM2	.77***	88.5	Substantial
	DM3	.76***	88.5	Substantial
Scope of discussion	S1	.70***	84.6	Substantial
	S2	.84***	92.3	Excellent
Learning processes	DLP	.68***	86.5	Substantial
	ILP	.89***	94.2	Excellent
Average		.78	89.6	

N.B. $N = 52$ for all codes. Statistical significance: ***<0.001, **<0.01, *<0.05.

The protocol therefore was established as a reliable tool. Discussion of these results follows at the end of the paper.

3. Teacher Learning Study

3.1. Methods

Having established a reliable analytic tool, the next part of our investigation concerned the use of the coding protocol for the systematic coding of the video data. As with the Instrument Development Study, the databases from Phases 4 and 5 were selected for this study for the reasons mentioned in section 2.1.2.

3.1.1. Participants

Four inclusion criteria were formulated for selecting schools from the two phases: (1) schools with at least two videos of at least six minutes each in a single LS phase; (2) schools with videos in which all teachers, or all teachers minus one[2], appeared on screen; (3) schools with videos of audio quality which was acceptable for coding; and (4) schools that had followed the project's LS procedure.

After applying these criteria, 31 schools from Phase 4 and 22 schools from Phase 5 remained. Fifteen schools were then randomly selected from each phase, giving a total of 30 schools (i.e. LS groups) to be sampled for analysis. Seven of the 30 schools participated in both the first and the second year of the project; the rest entered the project on the second year. Twenty-seven out of the 30 LS groups consisted of three teachers, two groups consisted of four teachers, and one group consisted of two

[2] *n-1* teachers present on screen would still mean that all teachers could be identified.

teachers, giving a total of 91 teachers (Table 7). Finally, 27 groups consisted of primary school teachers and three groups of secondary school teachers.

Table 7.
Number of teachers participating

Phase	Number of schools	Number of teachers in LS group	Total number of teachers
4	13	3	39
	1	4	4
	1	2	2
5	14	3	42
	1	4	4
			91

Teachers were asked to report their age, teaching experience, and sex, though not all complied. Thus, on the basis of the partial information provided, we can state that the estimated average age was 35.81 (*SD* =9.56; *n* = 57); the average amount of teacher experience was 9.32 years (*SD* =7.01; *n* = 31); 23.1% of the participants were male and 69.2% female (*n* = 84); 7.7% was missing data (*n* = 7).

In terms of materials used for the study, the reliable version of the coding protocol (Table 4) was used for the coding in the Teacher Learning Study. Samples from the videos provided by these 30 LS groups were analysed.

3.1.2. Procedure

From the videos of the 30 LS groups (or schools), four two-minute fragments, which will be called episodes, were randomly selected for each group. For this procedure, coders first randomly selected one video from a LS group. Then they randomly selected a minute of that video. That minute and the following minute made up one episode for coding. Coders repeated this process three more times for the same LS

group in order to end up with four episodes per LS group. This gave a total of 120 episodes to be coded.

As in the Instrument Development Study, a binary coding system was used, coding whether or not each code could be seen for each teacher separately. This procedure created a total of 364 data points as shown in Table 8. From these data points, 187 (51.4%) came from planning sessions and 177 (48.6%) came from reflective sessions.

Table 8.
Data points created from the 120 episodes

Phase	Number of schools	Number of teachers in LS group	Data points created from the selection of 4 episodes
4	13	3	156
	1	4	16
	1	2	8
5	14	3	168
	1	4	16
			364

The coded data was entered directly into SPSS. Having coded for all 120 episodes, the data was aggregated at teacher level and a value between zero and one was assigned for each teacher on each code. Since we had an equal number of observations per teacher, namely four observations, aggregation was possible. A value of 0 on a particular code meant that the teacher did not show that code in any of the four episodes. A score of .25 meant that the code was observed in one out of the four episodes; a score of .5 meant that the code was observed in two episodes and so on.

3.1.3. Analysis

We developed an analytical model based on our review of the literature and our protocol codes. Beyond a broad hypothesis that reflective sessions would produce more of our protocol codes than planning sessions, we theorized more specifically that teacher learning processes were likely to be influenced by being in a LS group (H1), in three respects. First, we expected that dialogic moves would facilitate teachers' learning processes (H2). Second, we expected teachers to show more and richer learning when they were talking about specific students (which are often the case students), since this enables them to share more specific observations and form more specific hypotheses for those students. Thus, we investigated whether talk about specific students has an effect on teacher learning processes and compared that to the effect of talk about groups of students (H3). Third, we expected that the two teacher learning processes were going to facilitate one another. More specifically, we expected that the more teachers in the LS group were engaged in DLP (descriptive learning), the more teachers would also engage in ILP (interpretative learning) (H4). These hypotheses are presented below:

H1: Variance in teachers' *individual* learning processes (both DLP and ILP) is partly located *at the group level*.

H2: Dialogic moves *at the group level* contribute to teacher learning processes *at the individual level.* Specifically:

> **H2a:** The level of questioning within the *LS group* contributes to teachers' individual learning processes.

H2b: The level of building on each other's ideas within the *LS group* contributes to teachers' individual learning processes.

H2c: The level of reasoning within the *LS group* contributes to teachers' individual learning processes

H3: The extent to which *LS group members* talk about specific students contributes to learning processes *at the individual level* more than the extent to which *LS group members* talk about groups of students.

H4: There is a positive relation between DLP *at the group level* and ILP *at the individual level*.

A two-level multilevel model was chosen as suitable for explaining the variability in teacher learning processes on both the individual and the group level. The statistical package HLM7 was used for our analysis (Raudenbush, Bryk, Cheong & Congdon, 2011). Assumptions of linearity and normality were met.

As represented in our hypotheses, the dependent variables (DVs) were the two teacher learning processes: DLP and ILP. A Pearson's correlation showed that the correlation between the two DVs was not significant. Therefore, instead of building a multivariate multilevel model, we decided to build separate models for the two DVs, meaning that hypotheses 1-3 were tested separately for each DV (Y_{ij}). The predictors in the models were the three dialogic moves for H2a-H2c, talk about particular students and talk about groups of students for H3, and DLP for H4. Data on these variables were aggregated into the group level and added to the models as second level predictors. Table 9 presents the descriptive statistics.

Table 9.
Descriptive Statistics of all variables at two levels

		N	Min	Max	Mean	*SD*
Level 1 variables	DM1	91	0	1	.51	.31
	DM2	91	0	1	.67	.26
	DM3	91	0	1	.43	.28
	S1	91	0	1	.61	.25
	S2	91	0	1	.52	.29
	DLP	91	0	1	.55	.28
	ILP	91	0	1	.53	.32
Level 2 variables	DM1	30	.17	.92	.51	.19
	DM2	30	.33	1	.68	.17
	DM3	30	.11	.67	.44	.13
	S1	30	.31	1	.62	.18
	S2	30	.13	1	.52	.24
	DLP	30	.17	1	.56	.19
	ILP	30	0	.92	.53	.23

The analysis consisted of the following steps:

i) H1 was tested with an intercept-only model, in which unique parts of the variance in teacher learning processes were allocated to the teacher and the LS group level.

ii) H2 was tested by adding the dialogic moves in teachers' talk on the group level (B_{1j}; B_{2j}; B_{3j}) to the model as predictors. Non-significant predictors were accordingly removed. We added dialogic moves on the individual level to this model, in order to test whether effects of dialogic moves on the LS group level would remain. Accordingly, all non-significant effects were removed from the model and the final model was tested.

iii) H3 was tested by building a model with talk about particular students (B_{4j}) and talk about groups of students (B_{5j}) on the group level as predictors. As with the second hypothesis, non-significant predictors were removed from the model.

iv) H4 was tested by using ILP as the DV and adding DLP on the group level (B_{6j}) as a predictor to the model.

v) Lastly, all significant predictors were integrated in one model. In the results, only significant predictors in the final model are reported.

Since the absolute values are not meaningful in themselves (i.e., they do not measure an absolute value), all group level variables were added (grand mean centred), and all individual level variables were added (group mean centred) to the model. In this way, we used group scores to compare between LS groups, and individual scores to compare between group members.

3.2. Results

3.2.1. Frequencies

Figure 4 below presents the frequency percentages for each code in each type of session.

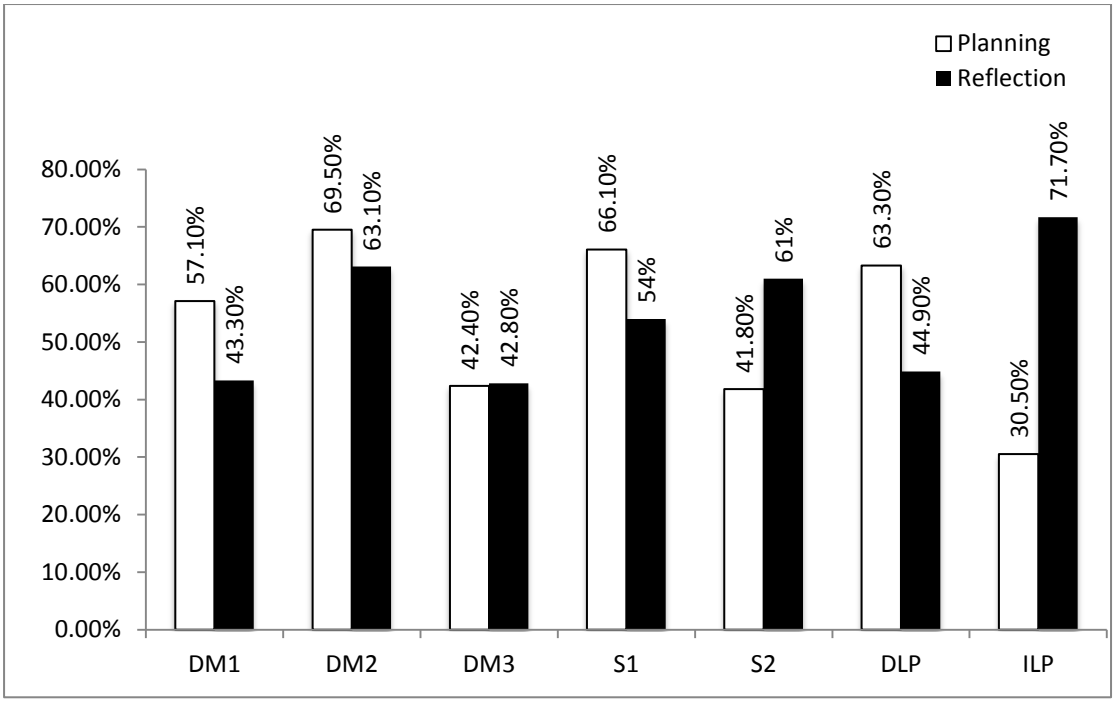

Figure 4. Frequency bar chart for planning and reflection episodes from final coding (120 episodes).

Our broad hypothesis (that reflective sessions would be likely to yield a greater number of our protocol codes than planning sessions) is disproved; there is, in fact, considerable variation. However, in reflective sessions over twice the number of ILP codes were assigned. This variance is considered in greater depth with respect to our specific hypotheses below.

3.2.2. Multilevel analysis results

3.2.2.1 Teacher learning processes at group and individual level

The results of the multilevel analysis confirmed our first hypothesis. A significant proportion of the variance in teachers' individual learning processes is located at the group level (Table 10), ICC= .21, χ^2 (29) = 52.314; p =.005, for teachers' DLP, and ICC= .29, χ^2 (29) = 64.313; p <.001, for teachers' ILP. Thus, respectively 21% of the

differences in teachers' DLP and 29% of differences in teachers' ILP are due to factors at the group level.

Table 10.
Results of the intercept-only model

	Descriptive learning processes		Interpretative learning processes	
Parameter	Estimate	S.E.	Estimate	S.E.
Group level variance	0.03	0.17	0.02	0.13
Individual level variance	0.07	0.27	0.06	0.25
Deviance	45.402514		28.212089	

3.2.2.2 Dialogic moves and teacher learning processes

The results for the dialogic moves and the scope of discussion are jointly presented in Table 11. The multilevel analysis for the dialogic moves revealed that the extent to which LS groups build on each other's ideas contributes positively to the extent to which individual teachers show DLP. Other dialogic moves on the group level did not significantly contribute to DLP on the individual level. Therefore, the second hypothesis – that the dialogic moves in the LS group would explain DLP at the individual level during LS meetings– is partly confirmed.

The coefficient of DM2 (i.e. building on each other's ideas) on group level could be interpreted as follows: Teachers in groups that show high levels (i.e. contributions of all group members in all four episodes) of DM2 are expected to engage in DLP in all episodes ($B_{0j} + B_{DM2}* 1>1$), whereas teachers in groups that show no DM2 at all are expected to engage in DLP in two out of four episodes ($B_{0j} + B_{DM2}* 0 = .55$).

With regard to H2c, the results show that the extent to which teachers provide evidence or reasoning – in comparison to their group members – contributes to the

extent to which they show DLP. A test was carried out to see whether these relations differ between groups; however, no significant random effects were found.

With regards to ILP, no significant contributions of dialogic moves were found. Hence, our second hypothesis – that dialogic moves in the LS group would explain ILP on the individual level – is rejected.

Table 11.
Results of multilevel analysis for dialogic moves and scope

	Descriptive learning processes		Interpretative learning processes	
	B (*SD*)	*T* (*d.f.*)	*B* (*SD*)	*T* (*d.f.*)
B_{0j}	.55(.03)***	19.76(27)	.53(.04)***	13.63(28)
DM2 on group level	.58(.14)***	4.17(27)		
S1 on group level			.50(.22)*	2.25(28)
S2 on group level	.22(.10)*	2.37(27)		
DM3 on individual level	.34(.08)*	2.50(60)		
S1 on individual level			.35(.16)*	2.22(59)
S2 on individual level			.35(.17)*	2.10(57)
Explained σ^2_e	48.06%		7.60%	
Explained σ^2_{u0}	16.09%		19.96%	
Deviance		8.928305		27.905106

Note: (a) Statistical significance: ***<0.001, **<0.01, *<0.05.

3.2.2.3. Scope of teacher discussions and teacher learning processes

The multilevel analysis revealed that the extent to which LS groups talked about particular students did contribute to the extent to which individual teachers showed DLP (Table 11). Accordingly, the third hypothesis, predicting that talking about particular students in the LS group would contribute more to DLP than talking about groups of students, is confirmed.

Furthermore, the results showed that the extent to which LS groups talked about groups of students did contribute to the extent to which individual teachers showed ILP. This explained 7.60% of the variance on the group level. Yet, this result is in contrast to our hypothesis that talking about particular students would contribute to more ILP than talking about groups of students.

Finally, the extent to which individual teachers talk about both particular students and groups of students contributes to the extent to which they show ILP. This explained 19.96% of the variance on the individual level. A test as to whether these coefficients differed across groups revealed no random effects.

3.2.2.4. The effect of descriptive learning processes on interpretative learning processes

No effect of DLP on ILP was found, suggesting that the extent to which LS groups showed DLP did not increase the chances for the particular group members to show more or less ILP. Our fourth hypothesis, which predicts a positive relation between DLP on the group level and ILP on the individual level, is therefore rejected.

3.2.2.5. Dialogic moves, scope of discussion and teacher learning processes

With regard to the variance in DLP, a substantial amount of variance at the group level is explained by both dialogic moves and the scope of the conversation. However, on the individual level, a lot of variance remains unexplained, while the difference of 36.47 suggests a good model fit. Furthermore, a chi-square test (χ^2 (1) =24.49, $p < .01$) showed that the final model does indeed explain DLP better than the empty model. We therefore conclude that both the quality of dialogic moves and the

scope of the conversation in a LS group make a substantial difference to the extent to which teachers' show DLP.

Consequently, the larger part of the variance in ILP, both at the group level and at the individual level, remained unexplained by dialogic moves and the scope of discussion. Of these two factors, only the scope of the conversation in a LS group explains, to some extent, whether teachers show ILP.

Lastly, we conclude that the two learning processes should be regarded as two separate processes; they do not influence each other and they are differently influenced by dialogic moves and scope. In other words, having group members that show a substantial amount of DLP does not increase the chances for an individual teacher in the group to show ILP.

4. Conclusions and Discussion

4.1. Research questions

This paper addressed two research questions with two separate, but inter-related, studies. These were:

- Can a reliable protocol be devised for analysing the relationship between dialogue and teacher learning processes?
- How is the content and structure of teachers' LS discussions related to the nature and quality of their learning?

4.2. Main findings and interpretations

In the context of a protocol with validity established through a bottom-up process of coding item analysis, the Instrument Development Study showed high reliability for

all seven codes of our protocol of analysing teachers' discussions. We therefore conclude that it is possible to devise a protocol for analysing the relationship between the content and structure of teacher's spoken dialogue and their learning processes. This protocol included an articulation of two fundamental teacher learning processes: descriptive learning processes, which have links to both teacher noticing (Sherin, Jacobs & Philipp, 2011) and reflection-on-action (Schön, 1987); and interpretative learning processes, which seem to occur primarily as teachers' widened discussion from individual case pupils to pedagogic consequences and intentions for wider groups.

In the Teacher Learning Study, the characteristics of LS teacher group discussions, and their contribution to teachers' individual learning processes, were investigated using the reliable coding protocol. The research question was addressed in the form of four hypotheses, which enabled the relationships between group level activity and individual learning processes to be evaluated (Section 3.1.3).

Further analysis was undertaken to assess the extent to which individual activity impacted on individual learning processes. This is a significant issue, as the extent to which individual activity in groups influences individual learning is still not well understood, even where students groups are the object of study (Howe, 2009). Interestingly, the difficulties of understanding this relationship in a naturalistic setting, such as a classroom, are hinted at in the structure of the PISA 2015 assessments (OECD, 2015), which recognise the importance of children acquiring skills in collaborative thinking but assess collaborative problem-solving skills individually, with a computer agent acting as a surrogate partner. In contrast, the research described here was carried out with teachers in 'conventional' working

groups. Furthermore, the findings reported here depend on the tight specificity of the rules of the coding protocol. Thus, for example, the way in which terms such as 'reasoning' and 'evaluation' are defined is subject to the need to ensure reliability within the coding protocol.

Our first hypothesis (H1), which predicted that the variance in teachers' individual learning processes (referring to both descriptive and interpretative learning processes) is partly located at the group level, was confirmed by the multilevel analysis. This is in line with previous research findings (e.g. Måseide, 2003). As discussed in Section 1.3, a positive framework for interaction within a professional group (indicated in our study by 'supportive moves' that enable the opening up of a dialogic space for professional interaction) enables learning.

The second hypothesis (H2, section 3.1.3) predicted that dialogic moves, including questioning (H2a), building on ideas (H2b) and reasoning (H2c), at the group level would contribute to both types of teacher learning processes (descriptive learning and interpretative learning) at the individual level. This hypothesis was partly confirmed. Whilst no effect was found on interpretative learning processes, an effect was found on descriptive learning processes by one of the three dialogic moves, namely 'building on ideas'. This 'building' was not always simply confirmatory or additional ('Yes…', 'And…', 'Also…'); there was also strong evidence of teachers drawing ideas together as they built on the preceding dialogue ('So…) and challenging statements by drawing upon additional knowledge ('But…', 'Although…'). The more teachers built on the ideas of others, the more they engaged in descriptive learning at the individual level. Since LS requires that teachers 'bring their knowledge to the

table', building on the ideas of others may mean that group members are better able to recall specifics and share details.

A further interesting finding was that the extent to which individual teachers providing evidence or reasoning contributed to the extent to which those individual teachers engaged in descriptive learning. It would seem that involvement in a Lesson Study group, with its associated 'ground rules' for engagement defined by the LS procedures, encourages participants to provide specific relevant knowledge, make explicit their reasoning and back that up with further evidence where necessary. This resonates with findings of research on student learning in groups (Howe & Abedin, 2013) and on what makes creative endeavour and problem solving most effective in adult working groups in a range of other contexts (Authors, 2013).

Interestingly, no direct relationship was found between dialogic moves and interpretative learning processes. Several explanations could be possible. The effect of dialogic moves in group interactions has been found not to have many observable effects in the short term on conceptual change. Howe (2009) has identified a delayed effect whereby dialogue promotes productive disagreement amongst group members, but may only lead to a change in ideas or understanding for its individual members after a significant passage of time. Thus, some of the effects of discussion in a LS group may take time to become evident in a teacher's talk and actions. If this is the case, it provides strong circumstantial support for the three-lesson cycle of LS, which gives time for teachers to interpret and evaluate their practice in the longer term.

The third hypothesis (H3, section 3.1.3) concerned the scope of discussion and predicted that the extent to which LS group members talk about specific students

would contribute to learning processes at the individual level more than the extent to which they talk about groups of students. This hypothesis was partly confirmed, as the results revealed that the extent to which LS groups talked about particular students contributed to how much individual teachers engaged in descriptive processes. However, what made a difference to the extent to which individual teachers engaged in interpretative learning was the extent to which the teacher *groups* talked together about groups of students. A possible explanation of this finding could be that descriptive learning is inextricably related to a focus on specific case students in the model of LS adopted in this study (Dudley, 2013). With this focus, descriptions shared in LS discussions are consequently largely based on detailed observations of particular pupils. Interpretative learning, on the other hand, is more likely to happen once teachers have reached common ground and generalize what they have found from specific students to groups of students. Interpretative learning therefore may abstract common principles from concrete cases and be more often related to pedagogic thinking about groups of students.

Testing the scope of conversation by *individual* teachers (that is, the extent to which individual teachers spoke about individual students or groups of students), it appeared that both talk about groups of students and talk about particular pupils was related to teachers' engagement in interpretative learning. A possible explanation could again lie in the concept of LS itself. Interpretations, in the form of evaluations and diagnosis in the present case, are a key part of teachers' professional learning through Lesson Study. Participating teachers were encouraged to be reflective, to adopt a 'researcher's perspective' while participating in the project and to employ specific dialogic approaches in order to do so. Interpretation and dialogue thus became part of the 'ethos of LS'.

Finally, the fourth hypothesis (H4, section 3.1.3), which predicted a positive relation between descriptive learning processes on the group level and interpretative learning processes on the individual level, was rejected. The fact that the two learning processes do not correlate within our LS groups resonates with learning qualities in student learning. As with deep and surface learning in students' learning (Authors, 2004b; Marton & Säljö, 1984), teachers can engage in descriptive and interpretative learning together or separately, in no particular order, to make sense of a situation.

Interestingly however, it was found that the more interpretative learning occurred at the group level, the more individual teachers engaged in this type of learning. This result can be explained by the findings in relation to the first hypothesis, which suggest that a significant proportion of the variance in teachers' individual learning processes is located at the group level. It is perhaps not surprising that the activities of a group affect the activities of its individual members.

To sum up, interpretative learning is more focused on connecting practice (concrete cases) to theory (abstractions), whilst descriptive learning involves teachers in learning at the concrete, practical level. We have shown that there are differential predictors of these two learning processes:

- When the group engages collectively in descriptive learning processes, in building on one another's ideas and in talking about individual students, then individual teachers show evidence of descriptive learning;

- When an individual teacher engages in reasoning, because they have to provide support for their statements they are more likely to show evidence of descriptive learning;

- When the group engages collectively in interpretative learning processes, and focuses on groups of students, then individual teachers show evidence of interpretative learning.

4.3. Implications for practice

Having developed a clear and reliable tool for the purposes of our research, we now have strong evidence that teachers' individual, professional learning can be positively influenced by being involved in genuine dialogue in a Lesson Study Group. The more specific findings have implications for teachers who are currently engaged or consider becoming engaged in LS practice. In particular, it is worth re-stating that the quality of a teacher's contributions to a group's discussions seem to be of paramount importance: building on the groups' shared contributions, and making strong individual contributions to reasoning, has an impact on teachers' descriptive learning processes. A final message that arises from the present study is that a positive ethos for collaborative group interaction, whereby members of a group provide supportive moves, is vital for learning processes to be achieved.

4.4. Suggestions for future research

This study makes a significant contribution to the field of teacher learning as it is one of the first to create and use a reliable coding protocol as a tool for investigating teachers' professional learning through spoken dialogue in a large-scale study. The protocol can be used as a tool for analysing dialogue in contexts for teachers' collaborative professional development other than LS. As the social norms of discussion can be culture-specific, however, a question that arises is whether the focus of our protocol, as well as the relationships between variables that we have been able

to detect, will have validity across cultural contexts. This is a matter for future research. Cross-disciplinary work with colleagues in Mathematics education is therefore a clear way in which further insights from this data may be drawn, and more widely indicates the value of cross-disciplinary teams in educational research.

Acknowledgements

This research was made possible through a grant of the London School Excellence Fund. We thank our colleagues at the Camden School Improvement Services and the teachers who participated in the study for their collaboration on the project.

References

Alexander, R. (2011). *Towards dialogic teaching: Rethinking classroom talk.* Cambridge: Dialogos.

Alexander, R. (2012). *Improving oracy and classroom talk in English schools: Achievements and challenges.* Retrieved October 15, 2013, from http://www.primaryreview.org.uk/downloads_/news/2012/02/2012_02_20DfE_ oracy_ Alexander.pdf.

Anderson, R. C., Chinn, C. A., Chang, J., Waggoner, M. A., & Yi, H. (1997). On the logical integrity of children's arguments. *Cognition and Instruction, 15*(2), 135-167.

Angelillo, C., Rogoff, B., & Chavajay, P. (2007). Examining shared endeavors by abstracting video coding schemes with fidelity to cases. In R. Goldman, R. Pea, B. Barron, & S. J. Derry (Eds.), *Video research in the learning sciences* (pp. 189-206). London: Routledge.

Bakhtin, M. M. (1981). In M. Holquist (Ed.), *The dialogic imagination: Four essays.* Austin & London: University of Texas Press.

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.

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education, 59*(5), 389-407.

Barnes, D., & Todd, F. (1977). *Communication and learning in small groups.* London: Routledge & Kegan Paul.

Borko, H., Jacobs, J., & Koellner, K. (2010). Contemporary approaches to teacher professional development. In E. Baker, B. McGaw, & P. Peterson (Eds.),

International encyclopedia of education (pp. 548-555). Oxford: Elsevier.

Bryk, A. S., Sebring, P. B., Allensworth, E., Luppescu, S., & Easton, J. Q. (2010). *Organizing schools for improvement: Lessons from Chicago.* Chicago, IL: University of Chicago Press.

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.

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.

Cochran-Smith, M. C. (2003). Learning and unlearning: The education of teacher educators. *Teaching and Teacher Education, 19*(1), 5-28.

Corcoran, D. (2011). Learning from lesson Study: Power distribution in a community of practice. In L. C. Hart, A. S. Alston, & A. Murata (Eds.), *Lesson Study research and practice in Mathematics education* (pp. 251-267). Dordrecht, Netherlands: Springer.

Dawes, L., Mercer, N., & Wegerif, R. (2003). *Thinking together: A programme of activities for developing thinking skills at KS2.* Birmingham: Questions.

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher, 38(*3), 181-199.

Dillenbourg, P. (1999). What do you mean by 'collaborative learning?'. In P. Dillenbourg (Ed.), *Collaborative-learning: Cognitive and computational approaches* (pp. 1-19). Oxford: Elsevier.

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.

Entwistle, N., & McCune, V. (2004). The conceptual bases of study strategy inventories. *Educational Psychology Review, 16*(4), 325-345.

Esser, J. K. (1998). Alive and well after 25 years: A review of groupthink research. *Organisational Behaviour and Human Decision Processes, 73*(2-3), 116-141.

Gijbels, D., Donche, V., Richardson, J. T. E., & Vermunt, J. D. (Eds.). (2014). *Learning patterns in higher education: Dimensions and research perspectives.* London: Routledge.

Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching, 8*(3), 381-391.

Hart, P. (1994). *Government: A study of small groups and policy failure.* Baltimore MD: The Johns Hopkins University Press.

Horn, I. S., & Kane, B. D. (2015). Opportunities for professional learning in mathematics teacher workgroup conversations: Relationships to instructional expertise. *The Journal of the Learning Sciences, 24*(3), 1-46.

Housley, W. (2000). Category work and knowledgeability within multidisciplinary team meetings. *Text, 20*(1), 83-107.

Howe, C. (2009). Collaborative group work in middle childhood. *Human Development, 52*(4), 215-239.

Howe, C., & Abedin, M. (2013). Classroom dialogue: A systematic review across four decades of research. *Cambridge Journal of Education, 43*(3), 325-356.

Isoda, M. (2015). The science of lesson study in the problem solving approach. In M. Inprasitha, M. Isoda, P. Wang-Iverson, & B. H. Yeap (Eds.), *Lesson study: Challenges in mathematics education* (pp. 81-108). Singapore: World Scientific.

Janis, I. (1982). *Groupthink: Psychological studies of policy decisions and fiascos* (2nd ed.). New York: Houghton Mifflin.

Jaworski, B. (2001). Developing mathematics teaching: Teachers, teacher educators, and researchers as co-learners. In F. L. Lin, & T. J. Cooney (Eds.), *Making sense of mathematics teacher education* (pp. 295-320). Dordrecht: Kluwer.

Jaworski, B. (2003). Research practice into/influencing mathematics teaching and learning development: Towards a theoretical framework based on co-learning partnerships. *Educational Studies in Mathematics, 54*(2-3), 249-282.

Jaworski, B. (2008). Building and sustaining inquiry communities in Mathematics teaching development: Teachers and didacticians in collaboration. In P. Sullivan, & T. Wood (Eds.), *The International Handbook of mathematics teacher education* (Vol. 3). Rotterdam: Sense Publishers.

Jaworski, B., & Huang, R. (2014). Teachers and didacticians: Key stakeholders in the process of developing mathematics teaching. *ZDM Mathematics Education, 46*(2), 173-188.

John-Steiner, V., & Mahn, H. (1996). Sociocultural approaches to learning and development: A vygotskian framework. *Educational Psychologist, 31*(3-4), 191-206.

Kamina, P., & Tinto, P. (2011). Lesson study: A case of the investigations mathematics curriculum with practising teachers at fifth grade. In L. C. Hart, A. S. Alston, & A. Murata (Eds.), *Lesson Study research and practice in Mathematics education* (pp. 221-233). Dordrecht, Netherlands: Springer.

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.* Dordrecht, Netherlands: Springer.

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

Lawrence, C. A., & Chong, W. H. (2010). Teacher collaborative learning through the lesson study: Identifying pathways for instructional success in a Singapore high school. *Asia Pacific Education Review, 11*(4), 565-572.

Lee, J. F. K. (2008). A Hong Kong case of lesson study - Benefits and concerns. *Teaching and Teacher Education, 24*(5), 1115-1124.

Lee, C. (2011). *Crossing borders: The global spread of lesson study.* University of Tokyo.

Lewis, J. M., Fischman, D., Riggs, I., & Wasserman, K. (2013). *Teacher learning in lesson study. Montana Mathematics Enthusiast, 10*(3), 583-620.

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

Linell, P. (2009). *Rethinking language, mind, and world dialogically: Interactional and contextual theories of human sense-making.* Charlotte, NC: Information Age Publishing.

Littleton, K., & Howe, C. (2010). *Educational dialogues: Understanding and promoting productive interaction.* Abingdon, Oxon: Routledge.

Littleton, K., & Mercer, N. (2013). *Interthinking: Putting talk to work.* Routledge.

Lonka, K., Olkinuora, E., & Mäkinen, J. (2004). Aspects and prospects of measuring studying and learning in higher education. *Educational Psychology Review, 16*(4), 301-323.

Marton, F., & Säljö, R. (1984). Approaches to learning. In F. Marton, D. Hounsell, & N. Entwistle (Eds.), *The experience of learning* (pp. 39-58). Edinburgh: Scottish Academic Press.

Måseide, P. (2003). Medical talk and moral order: Social interaction and collaborative clinical work. *Text, 2*(3), 369-403.

Matos, J. F., Powell, A., & Sztajn, P. (2009). Mathematics teachers' professional development: Processes of learning in and from practice (pp. 167-183). *The Professional Education and Development of Teachers of Mathematics: The 15th ICMI Study.* Dordrecht, Netherlands: Springer.

Mercer, N. (2000). *Words and minds: How we use language to think together.* London: Routledge.

Mercer, N. (2004). Sociocultural discourse analysis: Analysing classroom talk as a

social mode of thinking. *Journal of Applied Linguistics and Professional Practice, 1*(2), 137-168.

Mercer, N. (2008). The seeds of time: Why classroom dialogue needs a temporal analysis. *Journal of the Learning Sciences, 17*(1), 33-59.

Mercer, N., Dawes, L., Wegerif, R., & Sams, C. (2004). Reasoning as a scientist: Ways of helping students to use language to learn science. *British Educational Research Journal, 30*(3), 367-385.

Mercer, N., & Littleton, K. (2007). *Dialogue and the development of children's thinking*. London: Routledge.

Meyer, R. D., & Wilkerson, T. L. (2011). Lesson Study: The impact on teachers' knowledge for teaching mathematics. In L. C. Hart, A. S. Alston, & A. Murata (Eds.), *Lesson Study research and practice in Mathematics education* (pp. 15-26). Dordrecht, Netherlands: Springer.

Middup, C., Coughlan, T., & Johnson, P. (2010). How creative groups structure tasks through negotiating resources. In M. Lewkowicz, P. Hassanaly, M. Rohde, & V. Wulf (Eds.), *Proceedings of COOP 2010, computer supported cooperative work* (pp. 203-221). London: Springer-Verlag.

Murata, A., Bofferding, L., Pothen, B. E., & Taylor, M. W. (2012). Making connections among student learning, content, and teaching: Teacher talk paths in elementary mathematics Lesson Study. *Journal for Research in Mathematics Education, 43*(5), 616-650.

Myhill, D., Jones, S., & Hopper, R. (2006). *Talking, listening, learning: Effective talk in the primary classroom.* Maidenhead: Open University Press.

Nemeth, C., Rogers, J., & Brown, K. (2001). Devil's advocate v. authentic dissent: Stimulating quantity and quality. *European Journal of Educational Psychology, 31*(6), 707-720.

Niss, M., Peng Yee, L., & Kilpatrick, J. (2015). *Lesson Study: Challenges in mathematics education.* Singapore: World Scientific.

OECD. (2015). *Draft collaborative problem solving framework.* Retrieved February 15, 2016, from http://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Collaborative%20Problem%20Solving%20Framework%20.pdf.

Pintrich, P. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review, 16*(4), 385-408.

Postholm, M. B. (2012). Teachers' professional development: A theoretical review. *Educational Research, 54*(4), 405-429.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R. T., & du Toit, M. (2011). *HLM 7.* Lincolnwood, IL: Scientific Software International Inc.

Richardson, J. T. E. (2000). *Researching student learning.* Buckingham. SRHE and Open University Press.

Rojas-Drummond, S., Littleton, K., Hernandez,F., & Zúniga, M.(2010). Dialogical interactions among peers in collaborative writing contexts. In K. Littleton, & C. Howe (Eds.), *Educational Dialogues: Understanding and promoting productive interaction* (pp. 128-148). Abingdon, Oxon: Routledge.

Rowland, T., Huckstep, P., & Thwaites, A. (2005). Elementary teachers' mathematics subject knowledge: The knowledge quartet and the case of Naomi. *Journal of Mathematics Teacher Education, 8*(3), 255-281.

Runesson, U. (2008). Learning to design for learning: The potential of learning study to enhance teachers' and students' learning. In P. Sullivan, & T. Wood (Eds.), *The International Handbook of mathematics teacher education* (Vol. 1). Rotterdam: Sense Publishers.

Schön, D. A. (1987). *Educating the reflective practitioner.* Oxford: Jossey-Bass.

Sherin, M. G., Jacobs, V. R., & Philipp, R. A. (Eds.). (2011). *Mathematics teacher noticing. Seeing through the teachers' eyes.* New York: Routledge.

Shimizu, S., & Chino, K. (2015). History of Lesson Study to develop good practices in Japan. In M. Inprasitha, M. Isoda, P. Wang-Iverson, & B. H. Yeap (Eds.), *Lesson study: Challenges in mathematics education* (pp. 123-140). Singapore: World Scientific.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*(1), 1-22.

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

Stoll, L., Bolam, R., McMahon, A., Wallace, M., & Thomas, S. (2006). Professional learning communities: A review of the literature. *Journal of Educational Change, 7*(4), 221-258.

Tepylo, D. H., & Moss, J. (2011). Examining change in teacher mathematical knowledge through Lesson Study. In L. C. Hart, A. S. Alston, & A. Murata (Eds.), *Lesson Study research and practice in mathematics education* (pp. 59-77). Dordrecht, Netherlands: Springer.

Van Driel, J. H., Meirink, J. A., Van Veen, K., & Zwart, R. C. (2012). Current trends and missing links in studies on teacher professional development in science edu- cation: A review of design features and quality of research. *Studies in Science Education, 48*(2), 129-160.

Van Veen, K., Zwart, R., & Meirink, J. (2012). What makes teacher professional development effective? A literature review. In M. Kooy, & K. Van Veen (Eds.), *Teacher learning that matters: International perspectives* (pp. 3-21). New York, NY: Routledge.

Vermunt, J. D. (2013, 22 May). *Teacher learning and student learning: Are they related?*. Inaugural lecture given at the University of Cambridge, Faculty of Education.

Vermunt, J. D., & Vermetten, Y. J. (2004). Patterns in student learning: Relationships between learning strategies, conceptions of learning, and learning orientations. *Educational Psychology Review, 16*(4), 359-384.

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.

Vygotsky, L. (1962). *Thought and language*. Cambridge MA: MIT.

Webster-Wright, A. (2009). Reframing professional development through understanding authentic professional learning. *Review of Educational Research, 79*(2), 702-739.

Wegerif, R. (2007). Dialogic education and technology: Expanding the space of learning. *Computer-supported collaborative learning series.* New York: Springer Verlag.

Zeichner, K., & Liston, D. (1996). *Reflective Teaching: An introduction.* Mahwah, New Jersey: Lawrence Erlbaum Associates.

APPENDIX 1

Rules of coding

Code Definition	Rules for Coding	Examples from Data
[DM1] Requesting information, opinion or clarification	This code is assigned when a teacher asks: o Questions o Tag questions o Questions from the workbook	*If the plenary is to apply to different things, what would be a good thing to do?*
[DM2] Building on ideas	This code is assigned when a teacher responds to another teacher's statement with the following expressions: o Yes/Yeah/Hmhm/uhu (and)…. o And… o So… o But… o Although… o Also,…	*Yes, and I think that Ben, you hope, would now able to explain why both are a whole.*
[DM3] Providing evidence or reasoning	This code is assigned when a teacher is backing up an idea by saying 'because'.	*It's tricky because there's a spread of abilities and they have different learning intentions almost.*
[S1] Groups of pupils	This code is assigned when a teachers uses the following pronouns/expressions: o *They* o *Them* o *(This) group* o *The (least/most able) students* o *The top set/bottom set/middle group* o *People (referring to the students)* o *Pupils* o *Everybody (referring to the students)*	*I'd quite like to keep it like a word problem on the board when they are learning this backwards.*
[S2] Particular pupils	This code is assigned when teachers use the following pronouns or a student's	*I think she's made a lot of progress.*

	name: o He/Him/His o She/Her o [Name of student]	
[DLP] Descriptive learning processes	This code is assigned when teachers use a quote in their descriptions. This can be a quote of what a student said in the lesson or interview or a quote of proposed instructions for the future lesson. Code whenever you hear the following: o *He/she said….* o *You said…. (addressing the teacher who taught the lesson)* o *Then the students said….* o *He/she wrote…* o *Then you can say….*	*So you give them 4 pieces of an orange and say…no…you say' I've got 4 oranges and I want to share between 3 people. Each orange has 7 segments'. Or maybe not orange – something different*
[ILP] Interpretative learning processes	This code is assigned when observing evaluation of teaching, i.e. when teachers use words of evaluation such as the following: o Good o Nice o Effective o Fun o Clear o Easy o Difficult o Successful o Too (+adjective, such as difficult/quick/easy)	*I think that's why it's good. Because it's sticking with the same type of thing isn't it?*
	This code is also assigned when teachers' evaluate pupil learning, i.e. when they attach value to students' progress or compare students: o [Pupil A] found it too easy/hard/difficult o [Pupil B] found the answer/ was able to explain/ … o [Pupil A] did better than [Pupil B]	*I think she's made a lot of progress.*

	And finally this code is assigned when observing a teacher diagnosing, i.e. when a teacher identifies and/or classifies the type of problem in their pupils' understanding, knowledge or skills.	*There is something complex going on with him. He has learnt something else somewhere else or he has been taught without any visualisation.*