

University of Groningen

Teacher-Student Interaction Patterns Change During an Early Science Teaching Intervention

Menninga, A.; van Geert, P.; van Vondel, S.; Steenbeek, H.; van Dijk, M.

Published in:
Research in science education

DOI:
[10.1007/s11165-021-09997-3](https://doi.org/10.1007/s11165-021-09997-3)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2022

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Menninga, A., van Geert, P., van Vondel, S., Steenbeek, H., & van Dijk, M. (2022). Teacher-Student Interaction Patterns Change During an Early Science Teaching Intervention. *Research in science education*, 52, 1497–1523. <https://doi.org/10.1007/s11165-021-09997-3>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.



Teacher-Student Interaction Patterns Change During an Early Science Teaching Intervention

A. Menninga¹ · P. van Geert¹ · S. van Vondel¹ · H. Steenbeek¹ · M. van Dijk¹

Accepted: 16 February 2021 / Published online: 24 March 2021
© The Author(s) 2021

Abstract

This study aimed to explore the interaction between teachers and young students in terms of their question and answer patterns during science lessons and to investigate whether this changes over the course of an intervention called ‘Language as a Tool for learning science’ (LaT). It also compared experienced teachers with novices. A total of 16 teachers—of which 8 experienced and 8 novice teachers—and their students participated in this study. The teachers’ utterances were coded with regard to the use of questioning strategies, and the students’ reasoning skills were categorized by the use of three types of scientific reasoning skills (observations, predictions, explanations). Consistent with a complex dynamic systems orientation, we analyzed the interactions among these dimensions by means of state space grids (SSG). The results showed that before the intervention, the teacher-student interaction often took the form of a relatively rigid pattern in which teachers did not ask questions and students did not respond with reasoning expressions. In the course of the LaT intervention, a richer repertoire and a greater amount of interactions emerged in which knowledge was “co-constructed” by means of open-ended questions of the teacher and reasoning by the students. The results also suggest that the patterns of experienced and novice teachers were quite similar to each other.

Keywords Teacher-student interaction · Early science lessons · State space grid · Reasoning skills · Teacher questions · Video feedback coaching

Introduction

A myriad of components play a role in the development of students’ learning process and students’ general academic performance over time, but various studies have indicated that

✉ M. van Dijk
m.w.g.van.dijk@rug.nl

¹ Heymans Institute for Psychological Research, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands

especially the quality of the teacher-student interaction is of great significance (e.g., Downer et al. 2010; Geringer 2003; Hattie 2009; Sabol & Pianta, 2012). Teachers and students continuously and mutually influence each other in such a way that the teacher-student interaction at a certain moment in time is a product of a previous interaction and serves as the starting point for the next interaction (van de Pol et al. 2011; Steenbeek et al. 2012). For this reason, it has been argued that educational interventions should focus on improving the quality of the teacher-student interaction (Barber & Mourshed 2007). However, the evaluations of such interventions often do not provide information about the processes of change in the teacher-student interactions and the dynamics of these processes (Turner et al. 2014). As a consequence of this lack of knowledge about the processes of change, our understanding of the relation between real-time teacher behavior and associated student learning gains in interaction is still very limited. The current study aims at exploring how teachers and students interact in real-time during a series of science activities in terms of their questions and answers sequences. Specifically, we aim to investigate whether this interaction process changes over the course of an intervention called “Language as a Tool for learning science” (LaT) and whether experienced teachers are different from novices in this regard. In order to do so, we will employ a method of analysis—state space grids—that explicitly focuses on the processual, time-serial aspects of the interaction.

Currently, it is widely acknowledged that science education should be introduced at an early age (National Research Council 2000). In this context, the natural curiosity of young children is a good starting point, but learning can only take place when students are challenged to think and talk about science activities in an intellectually supportive environment (Gallas 1995; Lutz et al. 2006). The Framework for K-12 Science Education (National Research Council 2012) also stresses the importance of creating content-rich and discourse-rich classroom environments. These should include asking questions and defining problems, planning and carrying out investigations, interpreting observations, and formulating explanations.

A large body of literature exists on the issue of discourse and student-teacher interaction in science education, which has started with the seminal work of Sinclair and Coulthard (1975), Mehan (1979), and Cazden (2001) and the later contributions, specifically in science classrooms, of Chin (2006), Chin and Osborne (2010), Kelly (2014), Lemke (1990), Tan and Wong (2012), Van Booven (2015), Van Zee and Minstrell (1997), among others. It goes beyond the scope of this article to provide an overview of these approaches but it should be noted that the importance of teacher-student discourse in the classroom is universally recognized, and is considered crucial to scaffolding (Bruner 1986). One important structure of teaching discourse consists of a three-step sequence of teacher initiation, student response, and teacher evaluation (IRE) (Mehan 1979). Because the third step does not always include explicit evaluation, the sequence is also referred to as initiation, response, and follow-up (or feedback) (IRF) (Sinclair and Coulthard 1975). Wells (1993) has argued that when teachers use this structure effectively, this can be transformed into a cycle of learning and teaching that leads to co-construction of meaning. In a similar vein, Mortimer and Scott (2003) proposed to elaborate the sequence into an IRFRF chain, where the feedback from the teacher aims to evoke yet another response from the student and is followed up by further feedback. This can also be transformed into an explicit dialogic approach, according to these authors, whereas in traditional lessons, the purpose of questions is often to evaluate what students know; the nature of questioning in constructivist-based lessons is to encourage students to verbalize how they think and to help them construct conceptual knowledge.

The LaT intervention was developed at the department of Developmental Psychology of the University of Groningen in 2015. It is based on the above described insights from the literature on discourse on science education and the main goal of the program was to enhance and expand early elementary teachers' pedagogical-didactical skills, by improving questioning skills and linguistic skills that are required to verbalize students' reasoning processes. LaT was part of the national program "Curious Minds," which was initiated in 2006 by the Dutch Ministry of Education to study and stimulate science education in primary education. Specific to the LaT intervention is the focus on science education for young children (age 4–7 years) and the role of academic language. In the intervention, the empirical cycle (de Groot 1994) was introduced as an effective didactic means to structure the thinking process of students during science lessons. This means that teachers continuously use the sequence of asking research questions, predicting, testing, observing and analyzing, and drawing conclusions. The second goal was to make teachers aware of the important role of complex and sophisticated language during early science education. Language is essential to learning science and science lessons provide multiple language learning opportunities (Snow 2014; Wellington & Osborne 2001). During LaT, teachers are provided with information and strategies on modeling and evoking sophisticated language from students. One of the strategies to achieve this is to ask the questions from the empirical cycle in an open-ended manner. Open-ended questions are found to be most effective for stimulating children to talk (Oliveira 2010) and provide opportunities for more linguistically and cognitively challenging conversations (Massey et al. 2008). As a result of the improved science teaching skills of teachers of teaching science, students are expected to increase their scientific reasoning skills by verbalizing observations, predictions, and explanations. In the literature, many studies have indicated the positive impact of open-ended questioning strategies for stimulating and supporting students to talk (e.g., Massey et al. 2008; Oliveira 2010). Cognitive student gains of implementing these strategies are promising in (early) elementary school (Greenfield et al. 2009; Gelman & Brenneman 2004; Wetzels 2015). A recent evaluation of the LaT indicated that both teachers and their students benefitted from this intervention. Compared with a control group, the participating teachers used more open-ended questions, while students expressed more utterances related to reasoning. In addition, both teachers' and students' language use increased in syntactic complexity and—to a lesser extent—lexical sophistication (see Menninga et al. 2019). However, the main focus of the previous analyses was on either the behavior of teachers or the student outcomes—e.g., questioning skills, reasoning expressions, language complexity—and the question of whether the *interaction patterns* of teachers and students changed during the intervention has not been investigated yet.

The Framework of Complex Dynamic Systems

There is a growing consensus that the complex dynamic systems (CDS) approach provides a theoretical framework to understand interactions in real-time educational settings (Fischer & Bidell 2006; van Geert 1994, 2003). According to this framework, teacher-student interaction can be conceived of as *system* consisting of many interacting components that influence each other over nested time scales (Lewis 2002; Pianta 1999; Smith & Thelen 2003). These components are dynamic and transactional in nature. The system's behavior—what it "does"—is described by the sequence of its positions, or states, in a state space. A state space is formed by one or more (usually two) dimensions that are used to describe or characterize the system.

A key feature of a complex dynamic system is *self-organization*, which is spontaneous organization of the system towards a particular self-sustaining state, which is typical of the totality of influences and components that operate on a particular system (for which reason it can also be called the “preferred” state of the system, e.g., a particular school class and teacher). This self-sustaining state can be referred to as an *attractor state* characterized by recurrent patterns of interaction. Put differently, attractors are stable points, or regions, in the state space. Attractor states may differ considerably in terms of educational quality, that is, in terms of their effectiveness to contribute to specific goals of teaching and learning (Steenbeek & van Geert 2013). This attractor state is preferred by the system, but it does not refer to the quality of the interaction. An intervention, which can be seen as an external factor that sets a highly idiosyncratic process for individual teacher-student pairs in motion, is in fact a *perturbation* in an existing complex dynamic system (e.g., pattern of interaction). The aim of the perturbation is to change stable (but relatively ineffective) interaction patterns and replace them by new, more adequate, or desirable patterns that, once they are established, should also become self-sustaining (van Geert 2003). An example of a stable teacher-student interaction in a traditional teaching setting is that the teacher gives information and the students listen. A lot of energy is required in order to change such stable or rigid patterns of interaction into another, more optimal and more flexible way of interaction. An optimal interaction in terms of intervention goals (LaT) would be, for instance, that the teacher asks open-ended questions and the students actively use their reasoning skills.

Next to theoretical concepts, the complex dynamic systems approach offers methodological tools and techniques to capture the dynamics of change processes over time. A widely used example of such a tool is the state space grid analysis (SSG; Granic & Hollenstein 2003; Hollenstein 2007; Lewis et al. 1999). This methodology shifts the focus from averaged singular behaviors of either teachers or students to teacher-student *interaction sequence patterns*. The technique is often used in small sample studies for a more in-depth analysis of interaction processes. To begin with, it offers tools to visualize and analyze the quality of the interaction and its temporal structure. As stated above, the quality of the interaction is important in terms of behavior that is more optimal with regard to intervention goals. The structure of the interaction, which is the patterning of behavior (such as the interactional variability), informs us about the changing dynamics of the interaction patterns during the LaT intervention. Second, in addition to the qualitative possibilities, the SSG methodology offers various measures to quantify the quality of the interaction as well as its temporal structure. These measures can consequently be used for further quantitative analysis.

Using ordinal data of teachers and students, a “state space” can be constructed, which provides information about the moment-to-moment interaction over time, in terms of its quality and temporal structure. In other words, it reveals qualitative descriptions of the patterns of interaction in terms of coupled behaviors of teachers and students as well as general and specific measures of real-time variability in these patterns of interaction. Together, this shows when and how teacher and students coordinate their behavior. SSGs provide measures to determine the existence of a particular preferred state (short-term stability) and the pattern of preferred states over time (long-term stability). In educational practice, qualitative indicators of change in the system of teacher-student interaction can be (1) more coordinated interaction in terms of action-reaction patterns; (2) a change in preferred states of interaction; (3) the (increased or decreased) strength of the preferred states of interaction; and (4) the distribution of the types of interaction. In addition, indicators of change in the structure of interaction can be (5) changes in the variability in the types of interaction (Hollenstein 2013). Previous studies

have indicated that SSG can be effectively used to depict changes in patterns of interaction over time (Pennings et al. 2014; van Vondel et al. 2017). This suggests that the CDS theory and methodology offer a useful framework to understand the teacher-student interaction processes and their dynamic properties. The SSG methodology specifically enables us to describe changes in the moment-to-moment interaction patterns between teachers and students during a specific intervention, in this case the LaT intervention.

Teacher-Student Interaction and the Concept of Co-construction

The literature on scientific reasoning indicates that gains in student's reasoning skills can be measured by focusing on the predictions and explanations that students express (Henrichs & Leseman 2014; Treagust & Tsui 2014). In addition, students' descriptions of observations are an important starting point for moving to predictions and explanations (Fischer 1980). Hence, it is important for teachers to stimulate students' active participation, to spark students' curiosity, and to elicit students' thinking processes (Engel 2011). Teachers can use various instructional skills as means to make their classroom interaction more thought provoking with the aim of achieving deeper understanding in students. Although there is no one-size-fits-all approach in the educational practice (van de Pol et al. 2010), certain patterns of interaction (i.e., self-sustaining attractor states) may hinder active deep understanding, while others favor it (Ge & Land 2003; Granott 2005).

Scientific reasoning emerges in actual activities and verbal utterances of students while working on science activities guided by a teacher. It has been argued that the development of scientific reasoning can take place as a *co-construction process* (Sorsana 2008; Fischer & Bidell 2006), which means that together teacher and student create meaning. The term co-construction can refer to two things. First, it can refer to the *general process of mutual adaptation* creating any kind of outcome or pattern, called a "consensual frame" (Fogel 1993). An important feature of such processes is that teachers and students constantly adapt to each other's contribution during real-time interactions. Co-construction refers to the process of reciprocal influence, which in the case of classroom practices mainly concerns the behavior of a teacher that influences the reaction of a student and also vice versa (van Geert 1994). For instance, the co-construction process can lead to the desired learning outcomes in terms of the level of complex reasoning skills in real-time, but it can also lead to less desired superficial learning of declarative knowledge. The outcome typically depends on the nature of the self-organizing interaction patterns. An example of an interaction in which no co-construction of understanding occurs is a less desired and self-sustaining pattern of superficial teacher questions and instructions and superficial student answers on the declarative level. An interaction in which co-construction occurs might take the form of students expressing their thoughts and opinions and teachers posing open-ended questions (Oliveira 2010).

The second meaning is the co-construction of understanding, on which we shall focus in this article, is the *joint production or construction of a particular knowledge, insight, understanding*, etc. in the form of an iterative interaction between teacher and students, in a particular context of activity. The objective of the intervention is to move towards the construction of understanding in context and thus to move away from the recollection of facts (National Research Council 2013). Teachers can use various instructional skills as means to make their classroom interaction more thought-provoking with the aim of achieving deeper scientific understanding in their students (van Vondel et al. 2017).

Existing research indicates that there are large differences in the quality of teacher-student interactions in individual teachers at different moments in time (e.g., Geveke et al. 2017), but also differences between groups of teachers with different teaching experience. Experienced teachers seem to have the most stable, effective, and adaptive teaching skills, whereas novice teachers seem to have teaching skills that are less stable and effective, and consequently, they are much more capable of improving these skills (Clothfelter et al. 2007; Hattie 2003; Harris & Sass 2009; Heritage & Heritage 2013; Kane et al. 2008; Ladd 2008). The differences are largest when comparing teachers with less and those with more than 5 years of teaching experience (Clothfelter et al. 2007; Ladd 2008). One of the most notable differences between experienced and novice teachers concerns novices' issues with classroom management skills such as maintaining discipline and behavioral norms (Wolff et al. 2015). Co-construction processes require a lot of flexibility, which emerges from experience and knowledge of a great variety of classroom situations (Randi & Corno 2005). That is to say, co-construction processes take place in the form of typical attractor patterns (e.g., Geveke et al. 2017). A typical feature of such attractor patterns is the occurrence of self-sustaining sequences of open-ended questions by the teacher followed by open, exploratory, and divergent answers by the students, encouraging additional open-ended questions, and so on. Compared with experienced instructors, inexperienced teachers (novice teachers or pre-service teachers) have more limited frameworks of classroom management and interaction, making it harder for them to adapt to the abilities of their students. This often leads to teaching in a way that is more rigid and relies more on procedural structures (Sternberg & Horvath 1995) and lacks the flexibility to bring effective co-construction about.

In summary, many studies have stressed the importance of real-time teacher-student interactions and these can be conceptualized and analyzed as a complex dynamic system. This is also the case for science education, in which it is argued that teachers and students can co-construct science reasoning. However, our knowledge about the actual teaching-learning interaction processes in experienced teachers and novices is still limited. We also know only very little about how these processes in the case of experienced teachers and novices—and the possible differences between those groups—change in the long run.

Aims of the Study and Research Questions

The aim of this study was to explore the patterns of interaction between teachers and students in early grade science lessons and to investigate whether these change over the course of the LaT intervention and whether there are differences between teacher-student interactions of teachers with a different teaching experience (experienced teachers versus novices). We investigated the dynamics of the teacher-student interactions in terms of their sequential verbal behaviors. Here, we focused on open-ended questioning strategies for stimulating and supporting students to talk which have been demonstrated to have a positive impact on reasoning skills in students in the first years of primary education (Wetzels 2015). The analytic strategy was based on a complex dynamic system approach, with a focus on individual pathway, in this case individual classes, nonlinear shapes of change, and collective variables. As a first step, we explored real-time attractor patterns of teacher-student interaction and focused on whether and how teachers and students co-construct science discourse during the short-term timescale of a single lesson and whether and how this co-construction process changes over the course of the LaT (i.e., on the long-term timescale of the intervention). As a second step, we explored whether the processes of co-construction of teachers with different

teaching experience—that is experienced teachers versus novices—differed over the course of the intervention.

The following research questions were addressed:

RQ1. What are the patterns of experienced teachers and those of novices on the short-term timescale of interaction of teachers and students when co-constructing science discourse in terms of the quality (1a) and the structure (1b) of the interaction and how do these short-term patterns change over the course of the LaT intervention (long-term timescale)?

RQ2. Are there differences between experienced and novice teachers in the quality and the structure of the co-construction processes on the short-term timescale, and are there also differences in the long-term trajectory of these short-term timescale patterns?

Method

Participants

The data consisted of video recorded science lessons in naturalistic classroom situations. Teachers were recruited from schools in the North of the Netherlands. Eight experienced teachers and eight novice teachers participated in this study. They were recruited by means of flyers and personalized emails from schools in the North of the Netherlands. All participants had signed up on a voluntary basis. On average, the experienced teachers had 16 years of teaching experience (range 7–33), and were 41 years of age (range 30–60). They were all female. The novice teachers were all student teachers in the final year of teacher education who were enrolled in the bachelor program at the Hanze University of Applied Studies and who had followed a course in early elementary education. Seven novice teachers were female and one was male. The novices were on average 23 years old (range 21–26) and they had 4 years of teacher education experience each based on yearly internships. All teachers (both the novices and the experienced teachers) had limited experience in teaching science in early elementary school.

The participating teachers selected three to six students of their class, varying in age, gender, and cognitive level. The participating students (38 in total in the group of experienced teachers and 35 in total in the group of novice teachers) were around 5 years old (range 4–6). They were more or less evenly distributed according to gender (experienced teachers: 55% boys and 45% girls, novices: 49% boys and 51% girls). According to the teachers, none of the students they asked to participate had any notable developmental problems. All teachers and students were native speakers of Dutch. The teachers and parents of the participating students gave active informed consent before the start of the study, and all procedures conformed to existing ethical guidelines. This included written active consent was given by caretakers of children for both participation and video-recording for each participant. Participation was voluntary and no compensation was given. Data was pseudonymized and was stored following the guidelines of the Data Storage Protocol of the research institute. This study was approved by the local Ethical Committee Psychology.

Materials and Measures

The participating teachers were instructed to give a series of eight science lessons, with approximately one lesson each week. The choice of the lesson topic was left free in order to

support the teachers' self-efficacy. On request, teachers were given suggestions for the content of the lessons (for instance from the website www.proefjes.nl). Most teachers chose subjects such as floating and sinking, air pressure, designing, senses, and marble tracks. The first two lessons took place before the intervention started and will be referred to be the term "pre-measures." Immediately after lesson number 3 to 6, teachers received individual video feedback coaching in a post-class meeting (without the students being present). Two to 4 weeks after the final coaching session, teachers gave two further lessons, which were post-intervention measures (which we will refer to by the term "post-measures"). The total intervention period was 3–4 months. The intervention was adaptive in nature as the coaching was adjusted to the personal learning goals of the teachers, and to the particular situations from which the teacher-student interactions emerged. Measures were taken to ensure the quality of the implementation. First, the intervention was performed by an experienced coach (first author of this article). Secondly, only participants who completed the intervention were included, which is important with respect to the quantity of the intervention. One teacher dropped out after lesson 5 due to personal reasons. Third, the participant responsiveness, which is the degree to which the program stimulates the interest of participants, was considered substantial as all teachers invested their own time in preparing the lessons and joining the information meeting, and they all actively participated during the intervention.

For the analyses, 10 min (always the first 2.5 min, the middle 5 min, and the final 2.5 min) of each video recorded lesson were transcribed following the Codes for Human Analysis of Transcripts (CHAT) conventions (MacWhinney 2000) by the first author of this paper and a trained assistant-researcher. The group of students as a whole was taken as the unit of analysis, which means that the individual case is consisting of a group of individual students. In order to code the teacher and student utterances, the transcripts were exported to Excel. All off-task utterances that are unrelated to the task or materials at hand—such as "Teacher, can I go to the toilet?" or "Please go to your seat"—were removed before coding.

Teacher utterances were coded as (0) no question, (1) closed-ended question, or (2) open-ended question. The coding protocol was based on the coding scheme of De Rivera et al. (2005). The codes were mutually exclusive (see Table 1 for the descriptions and examples). The reasoning skills of the students were examined by tracking the use of three forms of scientific reasoning: observations, predictions, and explanations. Each task-related student

Table 1 Coding scheme for questions used by teachers including examples. Examples originate from the Dutch transcripts and are literal translations into English

Code	Type of expression	Example
No question (0)	Utterance related to instructions and information.	Please put the marble on the track. Keep your eyes on the table. Gravity is [...].
Closed-ended question (1)	Question that constraints the student's response because there is only one correct or acceptable answer, such as yes-/no-questions, multiple-choice questions, naming questions, et cetera.	Does it float? Will the rice also dance, or do you think it will stay at the bottom of the glass? So you say it will ...? What is this called?
Open-ended question (2)	Question that does not constrain the student's response because a number of different answers are acceptable (not one correct answer), such as open-ended questions, thought-provoking questions, elaborations questions, et cetera.	What just happened with the oil? What do you think will happen when [...]? Why do you think that [...]? How does that work then?

utterance was coded for (0) no expression of reasoning, (1) observation, (2) prediction, or (3) explanation. Table 2 provides an overview of the categories with examples of reasoning-related expressions of students. Observations (1) require the identification of salient characteristics of an object or event and can include perceptual or abstract relations. These observations are an important starting point for moving from initial observations to reasoning such as explanations. Predictions (2) require a mental representation of possible events that may happen in the future but have not happened yet. Explanations (3) result from deductive thinking about an outcome and demand causal relations in order to conclude about the underlying principles of a phenomenon. A composite measure of reasoning expressions was computed by adding up the absolute number of observations, predictions, and explanations and dividing this number by the total number of on-task utterances during the lesson.

Table 2 gives an overview of the categories with examples of reasoning-related verbal expressions of students, which were exhaustive and mutually exclusive.

The inter-observer agreement for the coding schemes was considered substantial (student reasoning: 78%) to almost perfect (teacher behavior: 88%). Cohen's kappa was calculated to determine the consistency of coding between the two observers (Landis & Koch 1977). The analysis revealed substantial consistency for both student reasoning ($\kappa = .61$) and teacher behavior ($\kappa = .79$).

Procedure

Participating teachers were instructed to give eight science lessons of 15 to 20 min within a period of 3 to 4 months with sessions every 1 or 2 weeks. Directly preceding the third lesson, the teachers attended an information meeting about the general principles of teaching science as formulated in the Curious Minds program (see authors). In this meeting, teachers were provided with information and tools on the empirical cycle, open-ended questioning strategies, scaffolding, and language learning strategies. Several video clips from the teachers' own pre-intervention measures were shown to illustrate the information previously given. After this, the teachers were asked to specify a personal learning goal that was used as a special point of interest for both the teacher and coach in the coaching sessions. This personal learning goal was aimed at stimulating the intrinsic motivation of the teachers and had to be in line with the coaching principles. An example of a teacher's personal learning goal was as follows: "I want to ask questions based on the empirical cycle" or "I want to reformulate students' language in more sophisticated wordings."

During the intervention-stage (lessons three to six), coaching was given immediately after every lesson. Coaching was based on the principles of video feedback coaching (Strathie et al. 2011; van den Heijkant et al. 2006) as this is an effective and useful method of (self-)reflection

Table 2 Coding scheme for reasoning of students including examples. Examples originate from the Dutch transcripts and are literal translations into English

Code	Type of expression	Example
No reasoning expression (0)	Task-related utterance, no reasoning.	To breath (answer to: why do we need air?).
Observation (1)	Student describes what he/she observes (sees, hears, feels, tastes, smells).	The marble sinks.
Prediction (2)	Student predicts what might happen in the future.	I think the balloon will make this sound [pffff].
Explanation (3)	Student gives reasons in order to explain phenomena.	Because it is so heavy.

which can be realized by entering the context of the teacher's actual teaching practice and by observing and discussing the teacher's behavior in real-time interactions with students (e.g., Domitrovitch, Gest, Gill, Jones & Sanford DeRousie, 2009; Seidel et al. 2011). During these sessions, the coach (also the first author) had selected several moments from the lesson that was just recorded, that included both moments with "optimal" interactions (e.g., the teacher using open-ended questions, the students showing reasoning skills, the students being actively involved) and moments that showed "less optimal" interactions (e.g., a teacher interrupting a student, students not responding to the teacher). The decision of what was "optimal" and what was "less optimal" was evaluated in terms of the teacher's personal learning goals and was determined on the spot by the coach. The coach always showed two "optimal" moments and "one less optimal" moment. Coach and teacher discussed and reflected upon these moments.

Data Analysis

State space grid (SSG) analysis was used for analyzing synchronized event sequences and as a visual illustration of teacher-student interaction (Hollenstein 2013). In general, each grid consists of the collections of all possible states of the teacher-student interaction. For data analysis, the teacher-student interaction sequence was taken as the unit of analysis. The unit of analysis is the teacher utterance and the subsequent student reaction. This is registered on the grid with a node. The succession of such nodes shows the trajectory of the teacher-student interactions, which concerns the short-time timescale of one lesson. Each node consists of an expression of the teacher (y -axis) and the following utterance of the students (x -axis).

The grid is divided into four qualitatively different quadrants of teacher-student interactions (see Fig. 1). The quadrant in the lower left (quadrant 1) represents *no co-construction*, a non-optimal interaction state in terms of the goals of the LaT intervention (the teacher asks no questions or closed-ended questions; students express no reasoning). By using the term "non-optimal," we do not mean to imply that these interactions are always non-optimal in an educational context. They are only non-optimal in relation to the explicit aim of the LaT to use open-ended questions to elicit reasoning. The quadrant in the upper right represents the most desired or optimal interaction state in terms of the intervention goals (quadrant 4): *active co-construction* (the teacher asks open-ended questions; students react with reasoning such as observation, predications, and explanations). The upper left and lower right quadrants (quadrant 2 and quadrant 3) represent interaction states in which a mismatch occurs between teacher and student input (in quadrant 2 teachers ask open-ended questions and students express no reasoning; in quadrant 3 teachers provide information/instruction or ask closed-ended questions and students reason). Each of these quadrants describes a potential attractor of student-teacher interactions (e.g., the majority of teacher and student interactions in a particular class and lesson might be confined to the no co-construction quadrant). SSGs were made for each pre-measurement, two intervention measurements (the second and the last coaching lessons), and each post-measurement per class.

The nodes in Fig. 1 depict the state of interaction at a certain moment in time. First, the interaction, represented by a node, is in the first quadrant (quadrant 1) of no co-construction. The arrow indicates a transition to another type of interaction, in this case to the fourth quadrant (quadrant 4) of active co-construction. The size of the nodes gives an indication of the time spent in a certain cell in the grid.

For further analysis, the SSG were, first, used for an exploration of how teachers and students co-construct science discourse over the course of an intervention (RQ 1). *Number of events* was calculated to gain insight into how many interactions were present during each

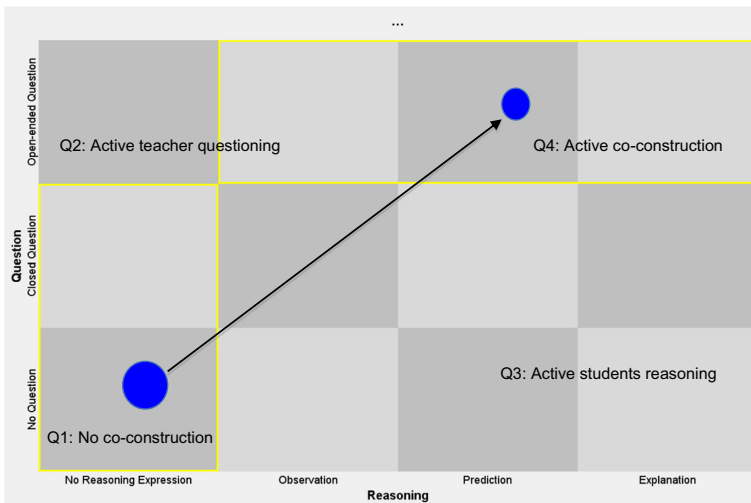


Fig. 1 The division of the four quadrants over the SSG. The y-axis represents the teacher behavior (questions) and the x-axis represents the student behavior (reasoning)

lesson. The *preferred state* was identified per lesson using the SSG measures “total events” and “events per cell.” The preferred state is operationalized as the quadrant in which most interaction occurs during that lesson, i.e., the cell with the most events. We calculated the percentage of interaction events within each quadrant, corrected for the total number of events within the grid, in order to study the *distribution of events* during each lesson. In addition, as we aimed to gain insight into the *strength* of these states, we used the SSG measurement called “return time.” This measure is based on the average duration of the intervals between visits to the selected cell and an indication of the strength of a preferred state (see Lewis et al. (1999)). A low return time is an indication of a relatively strong state, while a high return time is an indication of a relatively weak state. More information about the SSG measures can be found in Hollenstein (2013).

In order to be able to examine the degree of the system’s variability two dispersion measures were calculated. First, the *proportion of visited cells* was calculated to provide insight into the variability of the types of interaction (how many types of interaction occur during lessons). A low proportion of visited cells is an indication of a stable (sometimes rigid) pattern of interaction, while a high proportion of visited cells indicates more variability in types of expressed interaction. *Dispersion*, which is a whole grid measure, was calculated to indicate the number of visited cells controlled for the number of events per cell (how flexible the interaction is between the different types of interaction). Dispersion is expressed in a value between 0 (indicating no variability, all nodes are in a single cell) and 1 (maximum variability, there is an equal number of nodes in each cell). Again, to examine whether dispersion changed over the course of the intervention, the slope was tested against randomly shuffled data.

Next, SSGs were used for an exploration of the change in interaction patterns over the course of the intervention. This was done by calculating the slope per individual class over the six lessons. A Monte Carlo analysis (Todman & Dugard 2001) was employed to simulate the null hypothesis that the order of lessons is irrelevant; i.e., the probability of finding a similar or steeper slope is smaller than based on chance alone (procedural details, with examples, of this technique are provided in van Geert et al. (2011)). The proportions were randomly shuffled 10,000 times and each time the slope was calculated. The resulting *p*-values indicate the

probability that the slope of the empirical data would be found in the distribution of slopes of the shuffled data that represent instances of the null hypothesis. In addition, the effect size was calculated using Cohen's d (based on the means and the pooled standard deviations). Following Sullivan and Feinn (2012), an effect size of .20 or $-.20$ is considered small, .50 or $-.50$ medium, and .80 or $-.80$ large. Results will be interpreted using the p value and effect size: $p < .05$ and $d > .50$ or $-.50$ is strong evidence; $p < .1$ and $d > .50$ or $-.50$ is weak evidence; $p > .10$ and $d < .50$ or $-.50$ is considered no indication of change. For each group, an example of teacher-student interaction trajectory was chosen to illustrate the idiosyncratic process of change over the course of the intervention.

In order to be able to examine whether there are differences in the co-construction processes of experienced teachers and novice teachers (RQ2), the change trajectory of the different groups (as captured by the slope of each measure over the six lessons) were compared. A Monte Carlo analysis was now used to simulate the null hypothesis that both groups (i.e., experienced and novice teachers) came from the same population of teachers. This is done by randomly assigning every participating teacher, novice and experienced, to two groups only distinguished by their label (e.g., we call one group N and the other E.) For each of these random groups, we calculate the slope of the change trajectory. By repeating this random assignment a great number of times (10,000) we can determine the probability that the observed difference in slope occurs if novice and experienced teachers do not differ from one another in terms of the kind of change they undergo as a result of the intervention. Again, the results are interpreted in combination with the effect size. First, the pre-measurements were compared before the focus was on the change followed by a comparison of the changes over the course of the intervention.

Scatterplots were created to visualize how the interaction of individual participants changed from pre- to post-intervention. These scatterplots depicted both the differences within as well as between individuals and were informative on differences in the nature of change between the novice and experienced teachers. In the scatterplot, the percentages of the measures at pre-measurements are defined on the x -axis and the post-measurements on the y -axis. This scatterplot also contains the isocline, which is the line showing all scores for which no change occurs between pre-measurement and post-measurement. This line is constructed by creating dummy points consisting of every participant's pre-intervention score, and a post-intervention score similar to it. A real score falls above the isocline if it corresponds with positive change (increase) in for instance dispersion and below the isocline if it corresponds with negative change (decrease). The closer a score approaches the isocline, the smaller the pre-post-intervention change in the corresponding participant. By scrutinizing the distribution of the scores vis-à-vis the isocline, one can easily determine the distribution of the changes across the group, for instance, whether positive changes occur in the great majority of individuals, or whether an average difference between small sub-groups is due to a typical outlier.

Results

RQ1a: Experienced Teachers: Quality of the Interaction First, at pre-measurement, the average number of events for the experienced teachers was 32.8 (range 23–45). This means that there were around 33 instances in which verbal teacher behavior was followed by a student reaction, independent of the type of utterance. At post-measurements, this number was

32.3 (range 22–43). There was no change in the number of events over time (slope = $-.48$, $p = .80$, $d = -.32$).

The next step was to determine the quality of the interaction, based on the distribution of teacher-student interactions occurring in each quadrant (no co-construction, active teacher questioning, active student reasoning, active co-construction, see Table 3 (top)). In Fig. 2, four grids represent these results in a typical teacher-student interaction within lessons and over the course of the intervention. This case is representative of the group of experienced teachers and students in that the behavior of the system represented the interaction patterns of the majority of the experienced teachers. The selection of this specific figure is to some extent subjective and serves an illustrative purpose. That is, at pre-measurement, the interaction was predominantly in the “no co-construction” quadrant (Q1) (on average 67.3%, range 57–87%). This pattern was found for each teacher in the experienced teacher group. The percentage of interactions that emerged in quadrant 1 decreased over time ($M = 50.6%$, range 31–70%) and the probability that this decrease was based on chance alone was very low, providing strong support for change over time (slope = $-.04$, $p < .01$, $d = -1.13$). The preferred state of interaction at post-measurement was still in quadrant 1 for 6 out of 8 teachers. However, there was a decrease in the strength of the preferred state of interaction, indicated by a decrease in the average return time within this quadrant (slope = $.83$, $p = .05$, $d = 1.02$). This means that the interaction became more flexible over the course of the intervention.

At pre-measurement, the “active teacher questioning” quadrant (Q2) occurred in only 7.7% (range 4–24) of the interactions. There was strong evidence for an increase in the percentages of interactions ($M = 12.0%$, range 8–20%) that emerged in this quadrant (slope = $.01$, $p = .04$, $d = .70$). Interactions in the “active student reasoning” quadrant (Q3) occurred in 16.8% (range 7–31) at pre-measurement. The interaction within this quadrant is characterized by non-optimal teacher behavior in terms of the LaT intervention goals (giving instruction or information or posing closed-ended questions) and optimal student behavior (reasoning). There was no indication for a change over time ($M = 16.0$, range = 7–25%, slope = 0 , $p = .57$, $d = -.14$). At pre-measurement, 8.3% (range 0–19%) of the interactions occurred in the “active co-construction” quadrant (Q4). There was strong evidence for change over time; the percentage of interaction occurring in quadrant 4 increased ($M = 21.5$, range = 4–39, slope = $.03$, $p < .01$, $d = 1.03$). These findings show that a substantial part of the interactions moved away from non-optimal interaction patterns (Q1) towards more optimal interaction

Table 3 Distribution of teacher-student interaction occurred in each quadrant during pre-measurement, coaching session 2, coaching session 4, and post-measurement in the experienced and novice teacher group

	Quadrant 1 No co-construction	Quadrant 2 Active teacher questioning	Quadrant 3 Active student reasoning	Quadrant 4 DActive co- construction
Experienced teachers				
Pre-measurement	67.3%	7.7%	16.8%	8.2%
Coaching 2	55.8%	8.1%	16.7%	19.4%
Coaching 4	53.7%	8.7%	15.2%	22.5%
Post-measurement	50.6%	12.0%	16.0%	21.5%
Novices				
Pre-measurement	69.6%	8.8%	12.4%	9.2%
Coach 2	55.0%	5.7%	23.4%	16.0%
Coach 4	46.5%	12.3%	18.8%	22.4%
Post-measurement	46.6%	13.0%	21.5%	18.9%

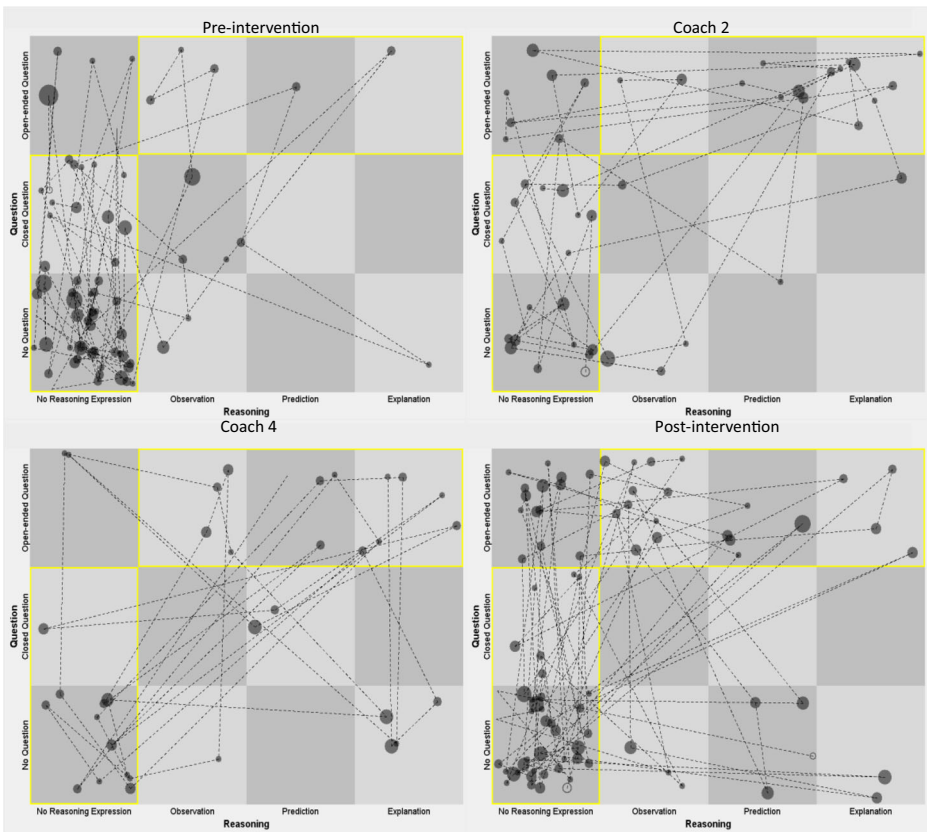


Fig. 2 Example of a microlevel state space grid of an experienced teacher. Teacher utterances are depicted on the y-axis using an ordinal scale of questions. Student utterances are depicted on the x-axis using an ordinal scale of scientific reasoning. Each grid represents a lesson (the first grid represents both pre-measurements, and the last grid represents both post-measurements). Each node represents the combination of teacher behavior and student reasoning expression

patterns (Q4) in the classroom in terms of the LaT intervention goals. At pre-measurement, a strong preferred state of interaction was quadrant 1 followed by the less strongly preferred state quadrant 3, while at post-measurement this second preferred state became quadrant 4.

RQ1b: Experienced Teachers: Structure of the Interaction The *proportion of visited cells* was, on average, .52 at pre-measurement, which indicates that about half of the cells in the grid were visited. There was no indication of change over time in the proportion of visited cells ($M_{pre} = .52$, $M_{post} = .59$, slope = .02, $p = .15$, $d = .54$). The calculated *grid dispersion* showed that there was an increase over the course of the intervention ($M_{pre} = .75$, $M_{post} = .83$, slope = .02, $p = .05$, $d = .74$). Therefore, there was strong support that there was more variability in interaction over time. This is illustrated by a typical case in Fig. 2. Taken together, the distribution and variability of the types of interaction changed while the number of visited cells remained the same.

RQ2a: Novice Teachers: Quality of the Interaction During the pre-measurements, on average, there were around 24.6 events in total (range 18–33) for the novice teachers. Over time, the number of events increased to 28.9 (range 20–37) at post-measurement. The probability

that the difference between pre- and post-measurement was based on chance alone was rather low (slope = .74, $p = .10$, $d = .75$). This supports our belief (weak evidence) that over the course of the intervention, there was slightly more interaction in the novice teacher group in terms of action-reaction patterns.

Table 3 (bottom) provides an overview of the distribution of interaction patterns over time. Figure 3 is an illustration of these results at the microlevel of one novice teacher within and over lessons. This case is typical of the group of novice teachers and students in that the behavior of the system was consistent with the interaction patterns of the majority of the novices. Again, this choice is to some degree selective and only serves the function of illustration. During the pre-measurements, the interaction pattern was predominantly in the “no co-construction” quadrant (quadrant 1) ($M = 69.6\%$, range 48–90%). For each of the novice teachers—and on average—this was the preferred state of interaction, which is considered non-optimal in terms of the LaT goals. The percentages of the different interaction patterns changed over time (post-intervention: $M = 46.6\%$, range = 31–68%). There was strong evidence for a decrease in the percentage of interactions in quadrant 1 in the course of the

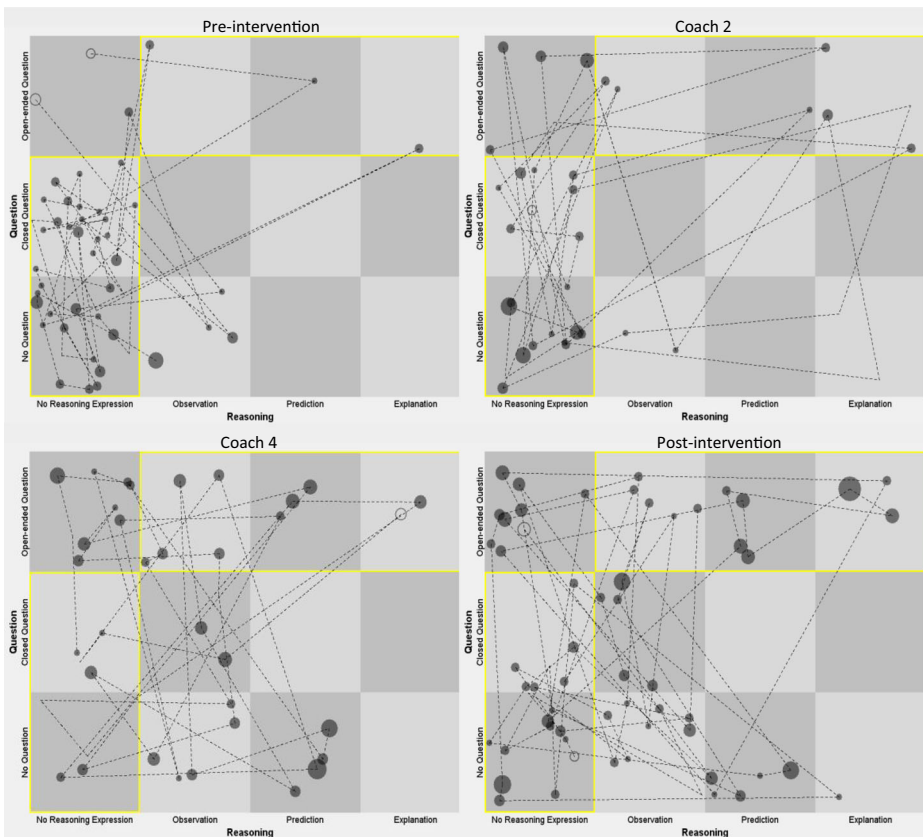


Fig. 3 Example of a microlevel state space grid of a novice teacher. Teacher utterances are depicted on the y -axis using an ordinal scale of questions. Student utterances are depicted on the x -axis using an ordinal scale of scientific reasoning. Each grid represents a lesson (the first grid represents both pre-measurements, and the last grid represents both post-measurements). Each node represents the combination of teacher behavior and student reasoning expression

intervention (slope = $-.05$, $p < .01$, $d = -1.73$). At post-measurement, quadrant 1 was still the preferred state of interaction (for 7 out of 8 teachers), but the interactions were less strongly concentrated in this state, as was indicated by a decrease in the mean return time towards this state (slope = $.87$, $p = .01$, $d = .71$). The “active teacher questioning” state (Q2) was visited in 8.8% (range = 4–13%) of the interactions. Over time, there was strong evidence for an increase in the percentage of interactions ($M = 13.0\%$, range = 0–24%) occurring in this quadrant (slope = $.01$, $p = .03$, $d = .67$). With regard to the quadrant of active student reasoning (Q3), 12.4% (range = 2–24%) of the interaction emerged in this quadrant at pre-measurement. There was strong evidence for change over time in the way that more interactions occurred in this quadrant ($M = 21.5\%$, range = 13–35%, slope = $.02$, $p = .03$, $d = .84$). Active co-construction (Q4) was observed in 9.2% (range = 0–25%) of the instances at pre-measurement. We found strong evidence for an increase in the percentage of interactions ($M = 18.9\%$, range = 6–32%) in quadrant 4 over time (slope = $.02$, $p = .01$, $d = 1.31$). These results showed that over the course of the intervention more open-ended teacher questions were followed by reasoning of students, which indicates more optimal co-construction. The distribution of the types of interaction changed over time, resulting in a different pattern of interaction in the classroom.

RQ2b: Novice Teachers: Structure of the Interaction At pre-measurement, the proportion of visited cells, on average, was .50. This means that approximately half of all possible cells in the grid were visited. We found strong evidence for an increase in the proportion of visited cells over time (slope = $.02$, $p = .03$, $d = 1.11$). This means that over time more cells were visited ($M_{\text{post}} = .64$), indicating more variability in the type of interactions. Secondly, the grid dispersion was calculated to give an indication of the variability of the interaction over the course of the intervention. We found strong evidence for an increase in this type of dispersion ($M_{\text{pre}} = .75$, $M_{\text{post}} = .87$, slope = $.03$, $p < .01$, $d = 1.54$). The illustration (see Fig. 3) shows that over time, the interaction becomes more scattered across the grid as the interaction emerges more often in all quadrants. Taken together, this means that there was more interaction and that the distribution of the types of interaction changed. In other words, the repertoire of the teacher-student interaction increased and the interaction became more flexible over time.

RQ3: Are There Differences in (the Dynamics of) the Co-construction Processes of Experienced Teachers and Novice Teachers? The second goal of the study was to explore whether there are differences in the co-construction processes in real-time teacher-student interactions between experienced teachers and novices. As a first step, the initial patterns of interaction were compared. An overview of the pre-measurement comparisons is provided in Table 4. Initially, there was a difference in the number of events between groups; there were

Table 4 Pre-measurement comparisons between experienced and novice teachers

	Experienced teachers	Novices	<i>p</i>	<i>d</i>
Events	32.8	24.6	< .01*	1.07
Q1 no co-construction	67.6	69.6	.36	-.14
Q2 active teacher questioning	7.7	8.8	.33	-.18
Q3 active student reasoning	16.8	12.4	.13	.36
Q4 active co-construction	8.3	9.2	.42	-.09
Proportion of visited cells	.52	.50	.40	.08
Dispersion	.75	.75	.45	.04

more action-reaction patterns within the experienced teacher group. We found strong evidence that the probability that this difference was based on chance alone was very low ($p < .01$, $d = 1.07$). The distribution of the types of interaction was very similar in both groups of teachers. There were no indications for differences regarding the percentages of interactions in quadrant 1 ($p = .36$, $d = -.14$), quadrant 2 ($p = .33$, $d = -.18$), Quadrant 3 ($p = .13$, $d = .36$), and quadrant 4 ($p = .42$, $d = -.09$) or proportion of visited cells ($p = .40$, $d = .08$) and dispersion ($p = .45$, $d = .04$). In sum, the co-construction processes seemed highly similar in both groups of teachers at the start of the intervention with one exception: the number of events was larger in the case of the experienced teachers.

The second step was to explore whether the changes in real-time teacher-student interactions were different for the experienced teachers and novices (for an overview see Table 5). We found weak support for a difference in change over time for the number of events ($p = .07$, $d = -.71$). This means that the extent to which the groups of teachers changed is somewhat different. The teacher-student interactions in the novice teacher group increased in the number of events, while this change was not found for the experienced teachers. At post-measurement, there was no longer an indication of a difference between the groups in the number of events ($events_{experienced} = 32.3$, $events_{novices} = 28.9$, $p = .13$, $d = .45$). With regard to the total number of events, Fig. 4 shows that in general there is a difference between the group of experienced teachers and novices, with some overlap of the lower range of the experienced teachers with the higher range of the novices.

Overall, the extent to which both groups of teachers changed in terms of the distribution of the interaction patterns was fairly similar. There were no indications of differences in the extent to which experienced teachers and novices changed over time with regard to the percentage of interactions in quadrant 1 ($p = .35$, $d = .19$), quadrant 2 ($p = .56$, $d = -.07$), and the quadrant 4 ($p = .22$, $d = .34$). We found weak support that the novice teachers changed more than the experienced teachers in the percentage of interactions in quadrant 3 ($p = .07$, $d = -.85$). However, Fig. 5 illustrates that the inter-individual differences are large. In the no co-construction scatterplot, it is expected that the majority of observations is below the isocline (decrease), whereas in the active co-construction scatterplot it is expected that the great majority should be above the isocline (increase). For the no co-construction (Q1) and the active co-construction (Q4) types of interaction, the differences between individuals— independent of their teaching experience— seem larger than the differences between the groups of experienced and novice teachers. The patterns of interaction greatly overlap in the pictures of no co-construction and active co-construction. Next, the focus was on comparing the structure of the interaction between experienced and novice teachers. There were also no

Table 5 Between-group comparisons of the average slope over time

	Experienced teachers Slope	Novices Slope	<i>p</i>	<i>d</i>
Events	-.48	.74	.07	-.71
Q1 no co-construction	-.04	-.05	.35	.19
Q2 active teacher questioning	.01	.01	.56	-.07
Q3 active student reasoning	-.002	.02	.07	-.85
Q4 active co-construction	.03	.02	.22	.34
Proportion of visited cells	.02	.02	.67	-.24
Dispersion	.02	.03	.72	-.31

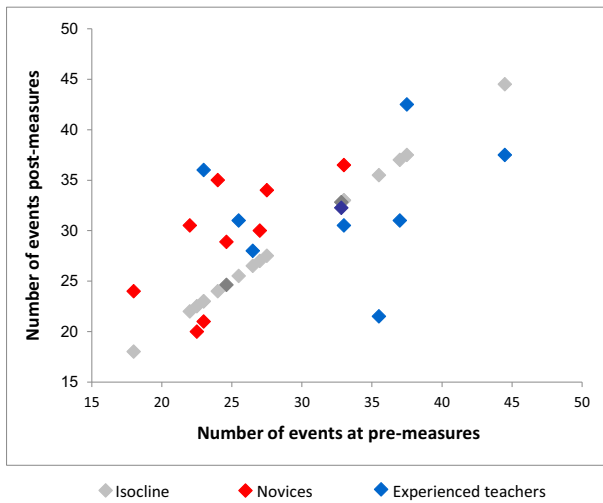


Fig. 4 Comparison of pre- and post-measurements of each individual for the total number of events. The gray diamonds \blacklozenge indicate the isocline based on the percentages in a state of interaction during pre-intervention measure. For the gray diamonds, the x-axis and y-axis both represent the number at pre-measurement, which indicates no change. The colored diamonds depict the number at post-measurement (\blacklozenge = novice teacher, \blacklozenge = experienced teacher). For the colored diamonds, the x-axis represents the pre-intervention number, and the y-axis the post-intervention number

indications of a difference between groups regarding the change over time in measures of variability (*proportion of visited cells*: $p = .67$, $d = -.24$; *grid dispersion*: $p = .72$, $d = -.31$). Figure 6 illustrates that the *proportion of visited cells* and *grid dispersion* demonstrated increases for almost all individuals. The dispersion measures of both groups of teachers greatly overlapped and teaching experience did not seem to be a clearly distinguishing factor looking at the dispersion values. However, there seem to be more novice teachers increasing in dispersion (i.e., in the way they shift between different types of interaction). Most experienced teachers, except for two, seem to stay around the isocline. This may point towards more rapid growth in the variability of the dynamics in the teacher-student interaction in the novice group.

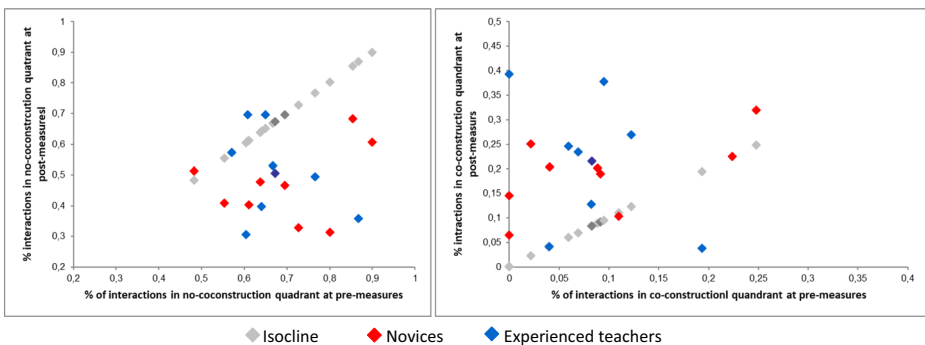


Fig. 5 Comparison of pre- and post-measurements of each individual for the percentage of interactions in Q1 (left) and Q4 (right) in the grid. The gray diamonds (\blacklozenge) indicate the so-called isocline based on the percentages in a state of interaction during pre-intervention measure. The colored diamonds depict the percentage at post-measurement (\blacklozenge = novice teacher, \blacklozenge = experienced teacher)

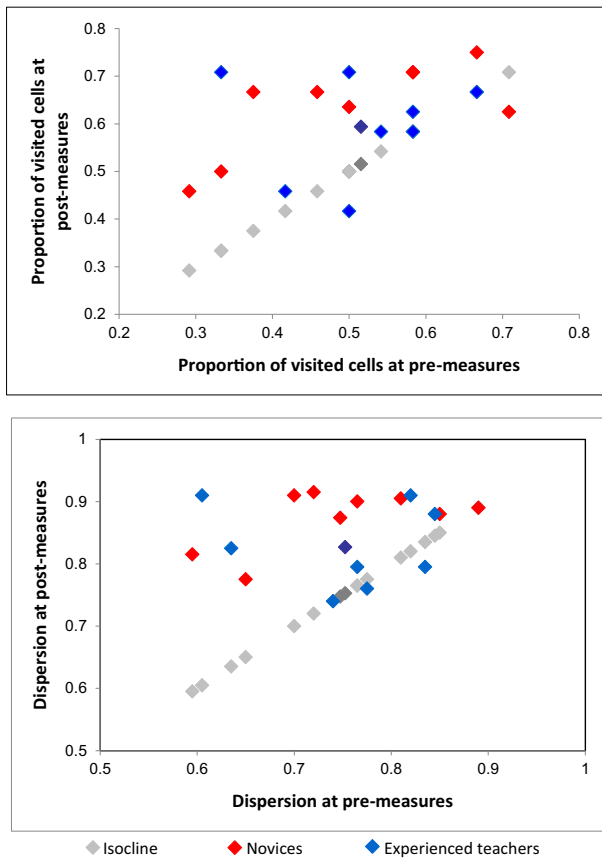


Fig. 6 Comparison of pre- and post-measurements of each individual for the measure of dispersion. The gray diamonds \blacklozenge indicate the so-called isocline based on the dispersion during pre-intervention measure. The colored diamonds depict the dispersion at post-measurement \blacklozenge = novice teacher, \blacklozenge = experienced teacher)

Discussion

Educational research stresses the importance of real-time teacher-student interactions, particularly when trying to optimize teaching outcomes by means of an intervention. However, our knowledge about the actual moment-to-moment teacher-student interaction processes before and during such interventions is still very limited. This is also the case for science education, in which it is argued that teachers and students co-construct science discourse and scientific reasoning. A study focusing on task-related real-time processes of the teacher-student interaction is therefore needed to enhance our understanding of learning and the changes that occur during an educational intervention.

This study aimed to explore the patterns of interactions between teachers and students in terms of quality and structure and investigated whether changes occur during the LaT intervention and whether there are differences between experienced teachers and novice teachers. In order to do this, the changes in real-time teacher-student interactions over time were analyzed by applying the method of state space grid analysis. This methodology shifts the focus from averaged singular behaviors of either teachers or students to teacher-student

interaction sequence patterns. The technique is often used in small sample studies for more in-depth analyses of interaction processes because it offers tools to visualize and analyze the quality of the interaction and its temporal structure. The quality of the interaction is important in terms of behavior that is more optimal with regard to intervention goals. The structure of the interaction, which is the patterning of behavior (such as the interactional variability), informs us about the changing dynamics of the interaction patterns during the LaT intervention. The SSG methodology offers various measures to quantify both the quality of the interaction as the structure. These measures can consequently be used for further quantitative analysis, as we have done in the present study.

The current study has provided insight into the dynamics of the real-time processes of change over the course of the LaT intervention. The results from the detailed microlevel analyses suggested that in the science lessons we observed before the intervention, the teacher-student interaction predominantly took the form of a relatively rigid pattern of interactions that consisted of closed questions or utterances without a question followed by students utterances that did not contain reasoning in which only limited co-construction took place. The system was drawn repeatedly towards this interaction pattern, in which it rested over extended periods and to which it returned quickly. This “no co-construction” quadrant was the preferred state before the intervention. The results also showed that during the LaT intervention a change in the patterns of interaction as well as to support the teacher-student system to employ a richer repertoire of teacher-student interactions was observed. The “active co-construction” quadrant, which is more optimal in terms of intervention goals, was visited increasingly often as the intervention proceeded. This second pattern consisted of open-ended questions that were followed by a reasoning expression of any of the students. The strength of the initial pattern decreased, which means that the behavior of teachers and students returned less quickly to this state. The finding that the teachers’ questioning strategies changed in the sense that they used more open-ended questions is in line with previous studies on the training of teachers in classroom settings (Wasik et al. 2006; Wolf et al. 1996), and also in the specific context of elementary science activities (Klein et al. 2000; Wetzels 2015). The increase in open-ended questions shows that the teachers made use of questioning strategies in their teaching practice, which may have contributed to a shift in teacher-student interaction from a mostly teacher-dominated style to a more student-oriented teaching practice. Moreover, the results of the interaction analyses showed that these are sequentially linked to the students’ reasoning expressions. This is in line with earlier studies that have shown that open-ended questions tailored to the abilities of students tend to elicit more reasoning (Chin 2006; Oliveira 2010; Newton & Newton 2000; van Vondel et al. 2017).

The results show that in this relatively short timeframe of the intervention, a new attractor pattern of less interactions of the type no co-construction and more active co-construction emerged. Although we cannot be sure that the specific features of the LaT intervention caused the changes in interaction dynamics we observed, the intervention was in fact intended to function as a perturbation in the sense that it aimed to actively change existing practices as a consequence of an external force (the coaching sessions). However, based on these results, we cannot conclude whether a pattern in which, for instance, the active co-construction pattern would quantitatively dominate over the no co-construction pattern would indeed be better in terms of learning effect than the current one. A balance between the different types of interaction—adapted to the specific (potential) abilities of the students and the learning environment—may be most beneficial, and this balance is most likely different for each teacher-student interaction at each moment in time.

With regard to the structure of the interaction, the results show that the variability of the interaction in terms of its sequential structure in general was rather high. High dispersion might be seen as a characteristic of adaptive classroom behavior (Geveke 2017). Over the course of the intervention, the interaction increased in variability, which reveals changes in the dynamics of the interaction. The interpretation of this increase in variability can be twofold. First, increasing variability can be an indication that the system became more erratic and disorderly. This is often seen as an indication of a process of change (Bassano & van Geert 2007; van Geert & van Dijk 2002). After a period of increased variability, the system can stabilize again. The results of the current study did not show this hypothesized decrease in variability, which might be an indication that the system did not yet stabilize into a new preferred state of interaction. Secondly, the increase in variability can also be an indication of more flexible and adaptive interaction between teachers and students. From a pedagogical-didactic point of view, this could mean that this fluctuation between preferred states within and over lessons may be interpreted as flexibility. Follow-up sessions in future intervention studies are needed to determine which interpretation applies to the current data. If the variability decreases, this may indicate stabilization of the system after a period of chaos (process of change). High variability in the long term may indicate an increase in flexibility in the teacher-student interaction.

The second aim of this study was to explore the differences between experienced and novice teachers. The results suggest that the interaction during science activities in both groups seems quite similar and independent of teaching experience. The extent to which both types of teachers benefitted from the intervention did not reveal differences at the group level. Visual inspection of scatterplots depicting the changes of individuals showed that the differences between individuals are large and that the two groups cannot be easily identified as different. The effect sizes of the novices were smaller than that of the experienced teachers. This may probably be caused by the fact that novice teachers show much more variability, implying that the standard deviation is bigger than with the experienced teachers. This directly affects the value of the effect size. The finding that teaching experience does not seem to play a major role in the quality of interaction during an intervention is not in line with previous studies (Clothfelter et al. 2007; Hattie 2003; Harris & Sass 2009; Heritage & Heritage 2013; Kane et al. 2008; Ladd 2008). A possible explanation for these results may be the fact that we only included the content-related verbal utterances. Previous research has shown that novice teachers are often more concerned with classroom management (Wolff et al. 2015). It may be that, in general, novices spent more time on interactions concerning management but that in terms of quality and content of the task-related interaction they are comparable to experienced teachers. The exploration of differences between experienced and novice teachers was particularly important in the light of developing and improving (existing) teacher professionalization.

The findings of this study have some important implications. First, the LaT had a strong focus on the observation of teacher-student interactions, and the results showed that during the intervention, teachers learned to recognize, and reflect upon, their own behavior, that of the students, and the complex interaction between both. We speculate that this combination of factors (video interaction guidance, reflection, and a focus on the system as a whole) was particularly important for establishing the observed behavioral change. Second, the results imply that the LaT intervention can be used for both experienced teachers and novice teachers, because the processes of development and learning and professionalization seem mostly similar. This means that the intervention—or the underlying principles of the intervention—

may be interwoven in the educational curriculum in teacher education. A third implication concerns the importance of within-lesson variability. The most notable change over time that could be observed was an increase in moment-to-moment variability in teacher-student interactions. This type of variability is also often seen as an indicator of change (Bassano & van Geert 2007; van Geert & van Dijk 2002) and, in this case, points towards the emergence of a greater repertoire in the interaction patterns. This stresses the relevance of including measures of variability for future research on intervention effectiveness, whether or not combined with more global measures, such as group averages of observed target behaviors.

This study has some limitations that should be taken into account in future studies. First, because of the lack of a control group for these analyses, we cannot be sure that the changes we observed were in fact *caused* by the LaT intervention. It may be speculated that these are the result of becoming more familiar with the science lesson content or with teaching in small teaching groups. More generally, although the results lead us to conclude that there is change over time with regard to teacher-student interaction, it is hard to attribute this change to the LaT intervention, or some other factor. However, as we have argued in Menninga et al. (2019), from a complex dynamic systems viewpoint, it is much less relevant to identify single causes for any change. Instead, change should be conceived of as resulting from a multi-causal process. For instance, it may be the case that type of science content may have had an influence on the interactions between teachers and students, and that some topics may be better suited to elicit teacher-student co-construction of understanding than others. Also, it is highly likely that the teachers' motivation and attitudes are particularly important in the context of early science education and also influence the teacher-student interaction. In this sense, the LaT intervention might merely create opportunities for changes in the teacher-student interaction. The causal mechanisms that create the changes in the interactions result from a complex interplay between a multitude of factors, in which not one thing can be singled out.

Secondly, the analyses reported in this paper are based on a small number of participants. There is always the inevitable trade-off between the number of participants and the depth of analysis. We need much more in-depth studies, each with probably a small number of participants in order to gain knowledge about the processes and ways to improve these teaching-learning processes in order for more optimal learning gains to establish. The in-depth study described in this paper is a first step.

Thirdly, the first author of this article was part of the team that developed the LaT intervention, was the coach, and coded the interactions together with a trained research assistant. This first author was not blind to the condition (novice vs experienced) and may subconsciously have influenced the results, both with regard to the coaching and the processing of the data. However, it should be noted that the trained independent inter-observer reliability on 20% of all transcripts was more than sufficient for all categories. At the same time, we should recognize that coaching is a dynamic process which is also influenced by many variables that are impossible to disentangle and which lead to highly idiosyncratic changes (Ridenour et al. 2013; Fisher et al. 2018).

Fourth, in this study, the “no co-construction” and “active co-construction” interactions were interpreted as non-optimal and optimal interactions in terms of the LaT intervention goals. Research on scaffolding, however, points out that there is no predetermined way—such as the ultimate one-size-fits-all-students approach—of stimulating reasoning in all students (e.g., van de Pol et al. 2010; van Geert & Steenbeek 2005). Van de Pol et al. suggest including the contingency of the teacher reaction. This means that the teacher's support is adapted to the current abilities of the students. This requires high flexibility of the teacher so that, given the

students' current level of functioning, the teacher responds in a way that fits with the current abilities of students and at the same time challenges the students to proceed to a higher level of functioning. In this context, the term “non-optimal” does not always apply to the situation. For instance, in the case of a teacher asking “Why do you think this happened?” and the student does not know what the explanation is or how to verbalize it, it may be that this question at this particular moment in time is out of the student's zone of proximal development (Vygotsky 1987). In this situation, the teacher's response should be contingent upon the student's reaction, which means that it lies within the zone of proximal development (Steenbeek et al. 2012). Giving cues or asking closed-ended questions can be of particular help to scaffold the student so that he or she can come up with an explanation. For future research, it is therefore suggested to include the contingency of the teacher behavior on the students' actions (van de Pol et al. 2011).

Another important suggestion for future studies is to focus on whole classroom interactions. Although small group activities, which were the focus of the current study, are increasingly becoming part of the everyday educational practice, the implementation of the LaT intervention strategies in a whole classroom setting calls for some (practical) adjustments. In the upper grades of primary education, a similar intervention yielded positive effects in a whole classroom setting (van Vondel et al. 2017). A starting point for exploring the teaching-learning processes in an early elementary whole classroom setting may be to do some follow-up sessions with the teachers who participated in the LaT intervention. These teachers have had the opportunity to explore, practice, and improve their science (and language) teaching skills. Some whole classroom follow-ups may provide insight in what form—in terms of quality and structure—the interaction patterns in these situations take. These insights are important for exploring the opportunities—and potential adjustments—to implement the intervention in whole classroom settings. The teacher-student interaction provides a unique entry point for educational interventions in that improving this interaction can be the direct focus of the intervention.

In conclusion, this study has demonstrated that in regular science lessons—before the intervention—the teacher-student interactions took place in a relatively rigid pattern of non-optimal interactions. This means that the system was drawn repeatedly towards a type of interaction characterized by the teacher asking closed-ended questions or not asking questions and students not expressing reasoning skills. This pattern remained over extended moments during the lessons, and the interactions returned to it quickly. The results further showed that the elements that were introduced during the intervention were powerful enough to change these patterns of interaction as well as to support the teacher-student system to employ a richer repertoire of teacher-student interactions. Over the course of the intervention, the interactions increased in variability, which reveals changes in the dynamics of the interactions. A comparison of the interaction patterns in experienced teachers and novices revealed that there are almost no differences in the content-related teaching-learning processes between both groups of teachers.

Funding This research was made possible with financial support from Platform Beta Techniek and the Heymans Institute for Psychological Research (University of Groningen, The Netherlands).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and

indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Barber, M., & Mourshed, M. (2007). *How the World's best education systems come out on top*. London, UK: McKinsey.
- Bassano, D., & Van Geert, P. (2007). Modeling continuity and discontinuity in utterance length: a quantitative approach to changes, transitions and intra-individual variability in early grammatical development. *Developmental Science*, 10(5), 588–612. <https://doi.org/10.1111/j.1467-7687.2006.00629.x>.
- Bruner, J. S. (1986). *Actual minds, possible worlds*. Cambridge, MA: Harvard University Press.
- Cazden, C. (2001). *Classroom discourse: the language of teaching and learning* (2nd ed.). Portsmouth, NH: Heinemann.
- Chin, C. (2006). Classroom interaction in science: teacher questioning and feedback to students' responses. *International Journal of Science Education*, 28, 1315–1346. <https://doi.org/10.1080/09500690600621100>.
- Chin, C., & Osborne, J. (2010). Supporting argumentation through students' questions: case studies in science classrooms. *Journal of the Learning Sciences*, 19(2), 230–284. <https://doi.org/10.1080/10508400903530036>.
- Clothfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). Teacher credentials and student achievement: longitudinal analysis with student fixed effects. *Economics of Education Review*, 26(6), 673–682. <https://doi.org/10.1016/j.econedurev.2007.10.002>.
- de Groot, A. D. (1994). *Methodologie: grondslagen van onderzoek en denken in de gedragswetenschappen*. s Gravenhage, The Netherlands: Mouton and Co..
- de Rivera, C., Girolametto, L., Greenberg, J., & Weitzman, E. (2005). Children's responses to educators' questions in day care play groups. *American Journal of Speech-Language Pathology*, 14, 14–26. [https://doi.org/10.1044/1058-0360\(2005/004\)](https://doi.org/10.1044/1058-0360(2005/004)).
- Downer, J., Sabol, T. J., & Hamre, B. (2010). Teacher–child interactions in the classroom: toward a theory of within- and cross-domain links to children's developmental outcomes. *Early Education and Development*, 21(5), 699–723. <https://doi.org/10.1080/10409289.2010.497453>.
- Engel, S. (2011). Children's need to know: curiosity in schools. *Harvard Educational Review*, 81(4), 625–645.
- Fischer, K. W. (1980). A theory of cognitive development: the control and construction of hierarchies of skills. *Psychological Review*, 87(6), 477–531. <https://doi.org/10.1037/0033-295X.87.6.477>.
- Fischer, K. W., & Bidell, T. R. (2006). Dynamic development of action and thought. In W. Damon & R. M. Lerner (Eds.), *Theoretical models of human development. Handbook of child psychology*. (pp. 313–399). New York, NY: John Wiley & Sons, Inc.
- Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences of the United States of America*, 115(27), E6106–E6115. <https://doi.org/10.1073/pnas.1711978115>.
- Fogel, A. (1993). *Developing through relationships: origin of communication, self and culture*. Chicago: The University of Chicago Press.
- Gallas, K. (1995). *Talking their way into science: hearing children's questions and theories, responding with curricula*. New York, NY: Teachers College Press.
- Ge, X., & Land, S. M. (2003). Scaffolding students' problem-solving processes in an ill-structured task using question prompts and peer interactions. *Educational Technology Research and Development*, 51(1), 21–38. <https://doi.org/10.1007/BF02504515>.
- Gelman, R., & Brenneman, K. (2004). Science learning pathways for young children. *Early Childhood Research Quarterly*, 19(1), 150–158. <https://doi.org/10.1016/j.ecresq.2004.01.009>.
- Geringer, J. (2003). Reflections on professional development: toward high-quality teaching and learning. *Phi Delta Kappan*, 84(5), 373–374. <https://doi.org/10.1177/003172170308400510>.
- Geveke, C. (2017). *It's not rocket science: developing pupils' science talent in out-of-school science education for Primary Schools* [Groningen]: Rijksuniversiteit Groningen.
- Geveke, C. H., Steenbeek, H. W., Doornenbal, J. M., & Van Geert, P. L. C. (2017). Attractor states in teaching and learning processes: a study of out-of-school science education. *Frontiers in Psychology*, 8, 299. <https://doi.org/10.3389/fpsyg.2017.00299>.

- Granic, I., & Hollenstein, T. (2003). Dynamic systems methods for models of developmental psychopathology. *Development and Psychopathology*, 15(3), 641–669. <https://doi.org/10.1017/S0954579403000324>.
- Granott, N. (2005). Scaffolding dynamically toward change: previous and new perspectives. *New Ideas in Psychology*, 23(3), 140–151. <https://doi.org/10.1016/j.newideapsych.2006.07.002>.
- Greenfield, D. B., Jirout, J., Dominguez, X., Greenberg, A., Maier, M., & Fuccillo, J. (2009). Science in the preschool classroom: a programmatic research agenda to improve science readiness. *Early Education and Development*, 20(2), 238–264. <https://doi.org/10.1080/10409280802595441>.
- Harris, D. N., & Sass, T. R. (2009). The effects of NBPTS-certified teachers on student achievement. *Journal of Policy Analysis and Management*, 28(1), 55–80. <https://doi.org/10.1002/pam.20402>.
- Hattie, J. (2003). Teachers make a difference, what is the research evidence? *Interpretations*, 36(2), 27–38.
- Hattie, J. (2009). The black box of tertiary assessment: an impending revolution. *Tertiary assessment & higher education student outcomes: Policy, practice & research*, 259–275.
- Henrichs, L. F., & Leseman, P. P. M. (2014). Early science instruction and academic language development can go hand in hand. The promising effects of a low-intensity teacher-focused intervention. *International Journal of Science Education*, 36(14), 2978–2995. <https://doi.org/10.1080/09500693.2014.948944>.
- Heritage, M., & Heritage, J. (2013). Teacher questioning: the epicenter of instruction and assessment. *Applied Measurement in Education*, 26(3), 176–190. <https://doi.org/10.1080/08957347.2013.793190>.
- Hollenstein, T. (2007). State space grids: analyzing dynamics across development. *International Journal of Behavioral Development*, 31(4), 384–396. <https://doi.org/10.1177/0165025407077765>.
- Hollenstein, T. (2013). *State space grids*. Springer US.
- Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education Review*, 27(6), 615–631. <https://doi.org/10.1016/j.econedurev.2007.05.005>.
- Kelly, G. (2014). Discourse practices in science learning and teaching. In N. G. Leder-man & S. K. Abell (Eds.), *Handbook of research on science education (Vol. 2)* (pp. 321–336). Mahwah, NJ: Lawrence Erlbaum.
- Klein, E., Hammrich, P. L., Bloom, S., & Ragins, A. (2000). Language development and science inquiry: the Head Start on Science and Communication program. *Early Childhood Research and Practice*, 2, 1–22.
- Ladd, H. F. (2008). Teacher effects: What do we know? *Teacher quality: Broadening and deepening the debate*, 3–26.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.
- Lemke, J. L. (1990). *Talking science: language, learning, and values*. Norwood, NJ: Ablex.
- Lewis, M. D. (2002). Interacting time scales in personality (and cognitive) development: intentions, emotions, and emergent forms. In N. Granott & J. Parziale (Eds.), *Microdevelopment: transition processes in development and learning* (pp. 183–212). New York, NY: Cambridge University Press.
- Lewis, M. D., Lamey, A. V., & Douglas, L. (1999). A new dynamic systems method for the analysis of early socioemotional development. *Developmental Science*, 2, 457–475.
- Lutz, S. L., Guthrie, J. T., & Davis, M. H. (2006). Scaffolding for engagement in elementary school reading instruction. *The Journal of Educational Research*, 100, 3–20. <https://doi.org/10.3200/JOER.100.1.3-20>.
- MacWhinney, B. (2000). *The CHILDES project: tools for analyzing talk. Volume I Transcription formats and programs* (Third ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Massey, S. L., Pence, K. L., Justice, L. M., & Bowles, R. P. (2008). Educators' use of cognitively challenging questions in economically disadvantaged preschool classroom contexts. *Early Education and Development*, 19(2), 340–360. <https://doi.org/10.1080/10409280801964119>.
- Mehan, H. (1979). *Learning lessons: social organization in the classroom*. Cambridge, MA: Harvard University Press.
- Menninga, A., van Dijk, M., Cox, R., Steenbeek, H., & van Geert, P. (2019). Co-Adaptation processes of syntactic complexity in real-Time kindergarten teacher-Student interactions. *Nonlinear dynamics, psychology, and life sciences*, 23, 229–260.
- Mortimer, E. F., & Scott, P. H. (2003). *Meaning making in secondary science classrooms*. Philadelphia, PA: Open University Press.
- National Research Council. (2000). *Eager to learn: educating our preschoolers* (p. 10.17226/9745). Washington, DC: The National Academies Press.
- National Research Council. (2012). *A framework for K-12 science education: practices, cross-cutting concepts, and core ideas*. Washington, DC: The National Academies Press.
- National Research Council. (2013). *Next generation science standards: by states, for states* (p. 10.17226/18290). Washington, DC: National Academies Press.
- Newton, D. P., & Newton, L. D. (2000). Do teachers support causal understanding through their discourse when teaching primary science? *British Educational Research Journal*, 26(5), 599–613. <https://doi.org/10.1080/713651580>.

- Oliveira, A. W. (2010). Improving teacher questioning in science inquiry discussions through professional development. *Journal of Research in Science Teaching*, 47(4), 422–453. <https://doi.org/10.1002/tea.20345>.
- Pennings, H. J., van Tartwijk, J., Wubbels, T., Claessens, L. C., van der Want, A. C., & Brekelmans, M. (2014). Real-time teacher–student interactions: a dynamic systems approach. *Teaching and Teacher Education*, 37, 183–193. <https://doi.org/10.1016/j.tate.2013.07.016>.
- Pianta, R. C. (1999). Enhancing relationships between children and teachers. *American Psychological Association*. <https://doi.org/10.1037/10314-000>.
- Randi, J., & Corno, L. (2005). Teaching and learner variation. Pedagogy: Teaching for learning. Monograph series II: Psychological aspects of education, 3, 47–69. <https://doi.org/10.1348/000709905X62110>.
- Ridenour, T. A., Pineo, T. Z., Molina, M. M. M., & Lich, K. H. (2013). Toward rigorous idiographic research in prevention science: comparison between three analytic strategies for testing preventive intervention in very small samples. *Prevention Science*, 14(3), 267–278.
- Sabol, T. J., & Pianta, R. C. (2012). Recent trends in research on teacher–child relationships. *Attachment & Human Development*, 14(3), 213–231. <https://doi.org/10.1080/14616734.2012.672262>.
- Seidel, T., Stürmer, K., Blomberg, G., Kobarg, M., & Schwindt, K. (2011). Teacher learning from analysis of videotaped classroom situations: does it make a difference whether teachers observe their own teaching or that of others? *Teaching and Teacher Education*, 27(2), 259–267. <https://doi.org/10.1016/j.tate.2010.08.009>.
- Sinclair, J., & Coulthard, M. (1975). Towards an analysis of discourse. London, UK: Oxford University Press.
- Smith, L. B., & Thelen, E. (2003). Development as a dynamic system. *Trends in Cognitive Sciences*, 7(8), 343–348. [https://doi.org/10.1016/S1364-6613\(03\)00156-6](https://doi.org/10.1016/S1364-6613(03)00156-6).
- Snow, C. E. (2014). *Deep comprehension: what does it look like in two- to three-year-olds?* [PowerPoint slides]. Retrieved from http://www.lucasacademie.nl/uimages/CatherineSnow_ECESS2014%20%282%29.pdf
- Sorsana, C. (2008). Introduction: developmental co-construction of cognition. *European Journal of Developmental Psychology*, 5(5), 537–543. <https://doi.org/10.1080/13546800701851179>.
- Sternberg, R. J., & Horvath, J. A. (1995). A prototype view of expert teaching. *Educational Researcher*, 24, 9–17. <https://doi.org/10.3102/0013189X024006009>.
- Steenbeek, H., & Van Geert, P. (2013). The emergence of learning-teaching trajectories in education: A complex dynamic systems approach. *Nonlinear dynamics, psychology, and life sciences*, 17, 233–267.
- Steenbeek, H., Jansen, L., & van Geert, P. (2012). Scaffolding dynamics and the emergence of problematic learning trajectories. *Learning and Individual Differences*, 22(1), 64–75.
- Strathie, S., Strathie, C., & Kennedy, H. (2011). Video enhanced reflective practice. In H. Kennedy, M. Landor, & L. Todd (Eds.), *Video interaction guidance, a relationship-based intervention to promote attunement, empathy and wellbeing* (pp. 170–180). London, U.K.: Jessica Kingsley Publishers.
- Tan, A.-L., & Wong, H.-M. (2012). ‘Didn’t Get Expected Answer, Rectify It’: teaching science content in an elementary science classroom using hands-on activities. *International Journal of Science Education*, 34(2), 197–222. <https://doi.org/10.1080/09500693.2011.565378>.
- Todman, J. B., & Dugard, P. (2001). *Single-case and small-n experimental designs: a practical guide to randomization tests*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Treagust, D. F., & Tsui, C. Y. (2014). General instructional methods and strategies. In N. G. Lederman & S. K. Abell (eds.) *Handbook of research on science education, Volume II* (303–320). New York, NY: Taylor & Francis.
- Turner, J. C., Christensen, A., Kackar-Cam, H. Z., Trucano, M., & Fulmer, S. M. (2014). Enhancing students’ engagement report of a 3-year intervention with middle school teachers. *American Educational Research Journal*, 51(6), 1195–1226. <https://doi.org/10.3102/0002831214532515>.
- Van Booven, C. D. (2015). Revisiting the authoritative–dialogic tension in inquiry-based elementary science teacher questioning. *International Journal of Science Education*, 37(8), 1182–1201. <https://doi.org/10.1080/09500693.2015.1023868>.
- van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in teacher–student interaction: a decade of research. *Educational Psychology Review*, 22(3), 271–296. <https://doi.org/10.1007/s10648-010-9127-6>.
- van de Pol, J., Volman, M., & Beishuizen, J. (2011). Patterns of contingent teaching in teacher–student interaction. *Learning and Instruction*, 21(1), 46–57. <https://doi.org/10.1016/j.learninstruc.2009.10.004>.
- van den Heijkant, C., Quak, G., van Swet, J., Vloet, K., De Vos, M., & van der Wegen, R. (2006). *School Video Interactie Begeleiding: Van meerdere kanten bekeken* (pp. 49–61). Apeldoorn, The Netherlands: Garant.
- Van Geert, P. L. C. (1994). *Dynamic systems of development: change between complexity and chaos*. London, UK: Harvester Wheatsheaf.
- Van Geert, P. L. C. (2003). Dynamic systems approaches and modeling of developmental processes. In J. Valsener & K. J. Connolly (Eds.), *Handbook of developmental psychology*. London, UK: Sage Publications Ltd..
- Van Geert, P. L. C., & Steenbeek, H. W. (2005). The dynamics of scaffolding. *New Ideas in Psychology*, 23(3), 115–128. <https://doi.org/10.1016/j.deapsych.2006.05.003>.

- Van Geert, P. & van Dijk, M., (2002). Focus on variability; new tools to study intra-individual variability in developmental data. *Infant Behavior and Development*, 25(4), 340–374. [https://doi.org/10.1016/S0163-6383\(02\)00140-6](https://doi.org/10.1016/S0163-6383(02)00140-6).
- van Geert, P., Steenbeek, H., & Kunnen, S. (2011). Monte Carlo techniques: statistical simulation for developmental data. In E. S. Kunnen (Ed.), *A dynamic systems approach to adolescent development* (pp. 43–52). London-New York: Psychology Press.
- van Vondel, S., Steenbeek, H., Van Dijk, M., & Van Geert, P. (2017). Ask, don't tell; A complex dynamic systems approach to improving science education by focusing on the co-construction of scientific understanding. *Teaching and Teacher Education*, 63, 243–253. <https://doi.org/10.1016/j.tate.2016.12.012>.
- Van Zee, E., & Minstrell, J. (1997). Using questioning to guide student thinking. *The Journal of the Learning Sciences*, 6(2), 227–269. https://doi.org/10.1207/s15327809jls0602_3.
- Vygotsky, L. S. (1987). Thinking and speech. In R. W. Rieber & A. S. Carton (Eds.), *L.S. Vygotsky, The collected works* (Vol. I, pp. 39–285). New York, NY: Plenum Press.
- Wasik, B. A., Bond, M. A., & Hindman, A. (2006). The effects of a language and literacy intervention on Head Start children and teachers. *Journal of Educational Psychology*, 98, 63–74. <https://doi.org/10.1037/0022-0663.98.1.63>.
- Wellington, J., & Osborne, J. (2001). Language and literacy in science education. Buckingham, UK: Open University Press. Retrieved from: <https://www.mheducation.co.uk/openup/chapters/0335205984.pdf>
- Wells, G. (1993). Reevaluating the IRF sequence: a proposal for the articulation of theories of activity and discourse for the analysis of teaching and learning in the classroom. *Linguistics and Education*, 5, 1–37. [https://doi.org/10.1016/S0898-5898\(05\)80001-4](https://doi.org/10.1016/S0898-5898(05)80001-4).
- Wetzels, A. F. M. (2015). *Curious minds in the classroom: the influence of video feedback coaching for teachers in science and technology lessons*. [Groningen]: University of Groningen.
- Wolf, S. A., Mieras, E. L., & Carey, A. A. (1996). What's after "What's that?": preservice teachers learning to ask literary questions. *Journal of Literacy Research*, 28(4), 459–497. <https://doi.org/10.1080/10862969609547937>.
- Wolff, C. E., van den Bogert, N., Jarodzka, H., & Boshuizen, H. P. (2015). Keeping an eye on learning: differences between expert and novice teachers' representations of classroom management events. *Journal of Teacher Education*, 66(1), 68–85. <https://doi.org/10.1177/0022487114549810>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.