



# Becoming a more adaptive teacher through collaborating in Lesson Study? Examining the influence of Lesson Study on teachers' adaptive teaching practices in mainstream secondary education

Tijmen M. Schipper<sup>a, b, \*</sup>, Rikkert M. van der Lans<sup>d</sup>, Siebrich de Vries<sup>b</sup>, Sui Lin Goei<sup>a, c</sup>, Klaas van Veen<sup>b</sup>

<sup>a</sup> Department of Human Movement and Education, Windesheim University of Applied Sciences, P.O. Box 10090, Campus 2-6, 8000 GB, Zwolle, the Netherlands

<sup>b</sup> Department of Teacher Education, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, the Netherlands

<sup>c</sup> LEARN! Research Institute, VU University Amsterdam, Van der Boechehorstraat 1, 1081 BT Amsterdam, the Netherlands

<sup>d</sup> Department Methodology and Statistics, Tilburg University, Warandelaan 2, 5037 AB Tilburg, the Netherlands

## HIGHLIGHTS

- Lesson Study plays a pivotal role in changing teachers' perceptions about adaptive teaching.
- Observation and self-report data can reinforce each other yet prove conflicting in other cases.
- The need for a clear and delineated definition for adaptive teaching remains.
- Measuring adaptive teaching behavior requires background knowledge of teachers' intentions and knowledge about students.

## ARTICLE INFO

### Article history:

Received 5 March 2019

Received in revised form

24 September 2019

Accepted 21 October 2019

Available online 12 November 2019

### Keywords:

Adaptive teaching

Lesson study

Classroom observation

Stimulated recall interviews

## ABSTRACT

Adaptive teaching has become increasingly important in research and practice. However, its complexity calls for effective professional development approaches, such as Lesson Study, that may promote adaptive teaching practices. This study uses a quasi-experimental mixed-methods design, consisting of stimulated recall interviews and classroom observation instruments, to determine whether participating in Lesson Study leads to more adaptive teaching practices in both teachers' perceptions and behavior. Although intervention group teachers reported various important changes in their perceptions about adaptive teaching and their (adaptive) teaching behavior, no intervention effects were found. This raises questions about how adaptive teaching can be defined and measured.

© 2019 Elsevier Ltd. All rights reserved.

## 1. Introduction

Adaptive teaching is a popular but complex construct (Corno, 2008; Deunk, Smale-Jacobse, De Boer, Doolaard, & Bosker, 2018; Parsons et al., 2018), and has increasingly received global attention due to inclusive education policies (UNESCO, 2009). In some countries, such as Canada or the Netherlands, integrating students with special educational needs in mainstream education is

promoted through national education policies (Jordan & McGhie-Richmond, 2014; Roy, Guay, & Valois, 2013; UNESCO, 2017). These policies may result in more heterogeneous classroom populations in terms of students' (cultural) backgrounds, motivation, abilities, and educational needs (Suprayogi, Valcke, & Godwin, 2017).

Adaptive teaching refers to addressing students' educational needs in the social and dynamic classroom context and entails both careful lesson planning and responsive teacher interventions in the lesson (Corno, 2008). Although the philosophy of addressing students' needs is hard to argue with (Hertberg-Davis, 2009), it confronts teachers with complex tensions between available time and willingness to differentiate their instruction, and the pressure of a high-stakes testing culture in many schools (Naraian &

\* Corresponding author. Department of Human Movement and Education, Windesheim University of Applied Sciences, P.O. Box 10090, Campus 2-6, 8000 GB, Zwolle, the Netherlands.

E-mail address: [t.schipper@windesheim.nl](mailto:t.schipper@windesheim.nl) (T.M. Schipper).

Schlessinger, 2018; Valli & Buese, 2007). In addition to this, teachers often lack confidence in their ability to implement inclusive practices (Dixon, Yssel, McConnell, & Hardin, 2014; Gehrke & Cocchiarella, 2013; Suprayogi et al., 2017; Wan, 2016), which might be a consequence of the complexity of adaptive teaching (Kyriakides, Creemers, & Antoniou, 2009; Van der Lans, Van de Grift, & Van Veen, 2018). Given the potential positive effects of adaptive teaching on student achievement and well-being (Deunk et al., 2018), the need for appropriate, effective and sustained teacher professional development (PD) activities that particularly promote adaptive teaching practices seems evident (Maeng & Bell, 2015; Suprayogi et al., 2017). Lesson Study (LS) offers unique opportunities to address this (Dudley, 2013).

LS is arguably one of the most rapidly growing PD approaches in the world (Dudley, 2015) and embodies many high quality features of PD (Perry & Lewis, 2009). Working with LS enables teachers “to develop the eyes to see children and how they respond and learn during research lessons” (Lee, 2015, p. 103). This can be promoted through the use of ‘case pupils’ (Dudley, 2013) who “represent or typify learner groups” of students (p. 110). LS is believed to support teachers to become more aware of students’ different educational needs, to feel more confident in addressing these needs and to be better able to adapt their classroom practices to these needs (Schipper, Goei, De Vries, & Van Veen, 2017; Norwich, Dudley, & Ylonen, 2014).

Measuring the effects of PD approaches such as LS on teachers’ adaptive teaching practices, however, has proven to be difficult (Deunk et al., 2018; Suprayogi et al., 2017). While observational data about actual teaching practices may strengthen research on adaptive teaching (Suprayogi et al., 2017), most studies in the context of LS research adopt qualitative designs with self-report measures (Xu & Pedder, 2015). Hence, only a few studies report the use of observation data, mostly in the context of research lessons or meetings as part of the LS cycle (e.g., Ni Shuilleabhain & Seery, 2018; Perry & Lewis, 2009; Warwick, Vrieki, Karlsen, Dudley, & Vermunt, 2019). However, to our knowledge, only one study uses observation instruments to capture the effects of LS on teachers’ daily adaptive teaching behavior in mainstream secondary education (Schipper et al., 2018).

Following this rationale, the aim of this study is to explore whether participating in LS leads to more adaptive teaching practices. Since adaptive teaching comprises more than actual teaching behavior in the classroom context (Beltramo, 2017), teachers’ intentions and perceptions about adaptive teaching are also included as previous studies found correlations between observation and self-report data. This is especially the case when the self-report data focused on “a teacher’s practices in a single class assignment and cover a clearly delineated and understood time frame” (Desimone, 2009, p. 189). In this study, both these measures are included using a quasi-experimental mixed-methods design consisting of two observation instruments, of which one was specifically constructed and validated for the purpose of this study, as well as stimulated recall interviews focusing on teachers’ adaptive teaching behavior in the observed post-test lessons. Adaptive teaching is further defined in the following section, showing the complexity of this construct.

## 2. Theoretical framework

### 2.1. Adaptive teaching behavior

Researchers generally refer to the term ‘differentiated instruction’ (Deunk et al., 2018) and its definition by Tomlinson and colleagues as “an approach to teaching in which teachers proactively modify curricula, teaching methods, resources, learning activities,

and student products to address the diverse needs of individual students and small groups of students to maximize the learning opportunities for each student in the classroom” (Tomlinson et al., 2003, p. 121). Various adaptations of this definition have subsequently been made, focusing on different elements to match students’ abilities such as the use of “systematic procedures for academic progress monitoring and data-based decision-making” (Roy et al., 2013, p. 1187). It is argued that “the core of differentiation is in teachers’ deliberate and adequate choices concerning instructional approaches and materials, based on well-considered goals and thorough analyses of students’ achievement, progress, and instructional needs, combined with continuous monitoring during the lesson” (Van Geel et al., 2019, pp. 10–11).

How, and to what extent, teachers decide to apply differentiated instruction “seems to be related to the implicit or explicit learning goals they have for their classroom as a whole” and may depend on whether teachers want to reach convergent goals (ensuring all students reach the same performance levels), divergent goals (ensuring that all students can reach their highest potential), or a combination of both (Deunk et al., 2018, p. 32).

Other researchers adhere to the term ‘adaptive teaching’ (Corno, 2008; Parsons et al., 2018) or ‘adaptive teaching competency’ (Vogt & Rogalla, 2009), of which the latter can be defined as teachers’ “ability to adjust planning and teaching to the individual learning processes of students” (Brühwiler & Blatchford, 2011, p. 98). Drawing on Corno’s (2008) concept of adaptive teaching, it is assumed that learning takes place within a socio-cultural context that requires adaptive teachers who value the diversity of talent in the collective. As such, adaptive teaching is not about individualized or individually tailored instruction where teachers adapt their instruction to “individual students in a social vacuum” (Corno, 2008, p. 165), but about teaching that is situated within the social context of the classroom. It is argued that adaptive teaching entails both teachers’ pre-planning efforts outside the classroom (i.e., macro-adaptations) together with “flexibly responding to students in the moment of teaching by improvising from previously established plans” (i.e., micro-adaptations) (Beltramo, 2017, p. 327).

As all these terms and definitions have a lot in common and seem to refer to an overall approach to teaching that implies careful monitoring of students’ progress and adaptive instruction accordingly, a clear focus is needed (Deunk et al., 2018; Suprayogi et al., 2017). In this study, we adhere to Corno (2008) rich description of adaptive teaching to stress that addressing students’ educational needs always takes place in a social and dynamic classroom context where, despite careful planning of the lesson, unexpected events occur that require flexible, spontaneous and responsive teacher interventions. The distinction between adaptive teaching and differentiated instruction tend to be rather subtle and may arguably be a more connotational issue as well. However, adaptive teaching seems to be a more broad approach to addressing students’ different educational needs and is generally associated with teaching in inclusive classroom contexts (Westwood, 2018), whereas differentiated instruction often refers to a more pragmatic, didactical approach focused on adopting specific teaching strategies and content (Suprayogi et al., 2017).

In addition to a certain adaptive mindset (Corno, 2008), adaptive teaching requires strong pedagogical content knowledge (Shulman, 1986), as well as a deep understanding of and familiarity with students to diagnose their needs in order to subsequently adapt their teaching methods, instruction and classroom management (Beltramo, 2017; Brühwiler & Blatchford, 2011; Vogt & Rogalla, 2009). Adaptive teachers can adjust their teaching to students’ needs in various ways such as through explaining, questioning, encouraging, challenging, coaching, feedback, modeling, and (formative) assessing (Parsons et al., 2018; Smit & Humpert,

2012), and by focusing on four curriculum-related elements: the subject matter content, the learning process, the product of learning, and affective elements in the learning environment (Tomlinson & Imbeau, 2010). When referring to students' needs, one can distinguish students' readiness, interest, and learning profile, as well as students' characteristics such as (cultural) background, self-awareness, confidence, and independence (Tomlinson et al., 2003; Tomlinson & Imbeau, 2010). To demarcate the focus in this study, we specifically refer to students' learning needs in line with Deunk et al. (2018).

Based on this theoretical outline, we define adaptive teaching as carefully and proactively planning of the curriculum, teaching materials and learning activities, as well as flexibly responding to students learning needs in the social context of the classroom in order to reach the desired lesson objectives. We argue that adaptive teaching can be categorized into teacher-student interactions and adjusting the content to students' learning needs. The first may contain elements such as providing support and encouragement to individual students, as well as clear feedback and recommending appropriate learning strategies. The latter may relate to how teachers differentiate the lesson tasks, work formats, and how they assign homework, but also on what level teachers explain the content.

## 2.2. Lesson Study as a means to foster adaptive teaching behavior

The popularity of the Japanese teacher PD approach LS is rapidly increasing around the globe in both research and practice (Dudley, 2015). Despite its various (cultural) adaptations (Stigler & Hiebert, 2016) and the influence of the school context in which it is implemented (Schipper et al. 2017, Schipper, De Vries, Goei, & Van Veen, 2019; Hadfield & Jopling, 2016), LS shares several essential features and involves small groups of teachers who collaboratively identify a research focus, study the curriculum, design a research lesson in detail, teach and observe the research lesson, and thoroughly evaluate the research lesson in a post-lesson discussion (Godfrey, Seleznyov, Anders, Wollaston, & Barrera-Pedemonte, 2019; Lewis, Perry, & Murata, 2006). A LS cycle is often guided by a LS facilitator or an outside specialist who serves as a knowledgeable other (Bjuland & Helgevol, 2018; Takahashi & McDougal, 2016).

Teachers who are part of a LS team can combine academic and social-behavioral goals in their research focus, but a LS cycle is generally focused around traditional curriculum subjects, mathematics in particular (Norwich, Fujita, Adlam, Milton, & Edwards-Jones, 2018). Although LS often leads to developed lesson plans and material at the end of a LS cycle, this is not the objective. The main purpose of LS is to gain more knowledge about and insight into (the learning of) students as well as to improve the quality of didactical decisions and lesson design (Wake, Swan, & Foster, 2016). One of the main reasons for its popularity is that LS offers unique possibilities to closely observe student learning (Dudley, 2015). The considerable time and thought that is devoted in LS to predicting how students might respond to different teaching situations and material (Dudley, 2013; Norwich & Ylonen, 2013), can be transferred to teachers' daily practice and may enable teachers to develop greater awareness of students' diverse learning needs and to develop adaptive teaching strategies in turn (Lee Bae, Hayes, Seitz, O'Connor, & DiStefano, 2016).

While the evidence base for LS is growing at a high pace, the positive impact that participating in LS can have on teachers' knowledge and behavior, teacher collaboration and professionalism, as well as student learning, becomes more visible (Huang & Shimizu, 2016; Xu & Pedder, 2015). As discussed, most of these results stem from self-report studies though there is also evidence

from randomized controlled trial experiments (Lewis & Perry, 2017) and quasi-experimental research (Schipper et al. 2018; Helgevol, Næsheim-Bjørkvik, & Østrem, 2015), focusing on different outcome variables and contexts. The majority of LS research is situated in primary and secondary education, and is generally focused around specific content matter (Xu & Pedder, 2015). However, LS can also be applied in multidisciplinary contexts (Schipper et al. 2017) and receives increasing attention in initial teacher education (Leavy & Hourigan, 2016).

In the context of adaptive teaching, LS research has mainly focused on students with (moderate) learning difficulties (Norwich et al., 2018; Norwich & Ylonen, 2013) and only to a little extent in mainstream secondary education (Schipper et al. 2017; Schipper et al. 2018). Consequently, with this study we aim to increase the knowledge base about adaptive teaching in the context of mainstream secondary education by combining both self-report and observation measures. While self-report studies could provide more insight into adaptive teaching practices in terms of teachers' attitudes, beliefs or adaptations in their curriculum (Beltramo, 2017; Dixon et al., 2014; Suprayogi et al., 2017; VanTassel-Baska & Stambaugh, 2005; Wan, 2016), "without additional data based on classroom observation, it is difficult to know whether teachers' perceptions are indeed echoed in practice" (Bruggink, Goei, & Koot, 2016, p. 8) and how these practices may be supported by LS.

## 2.3. Measuring adaptive teaching behavior

Capturing adaptive teaching behavior has proven to be difficult, "especially when fuzzy constructs like 'differentiation' are the topic of concern" (Deunk et al., 2018, p. 32). Various observation instruments have been developed over time, each with a different focus, sometimes including adaptive teaching items (Bell, Dobbelaer, Klette, & Visscher, 2019; Pianta & Hamre, 2009; Schoenfeld, 2013). However, most of these studies focus on generic teaching behavior (Bell et al., 2019) or a particular group of students such as students with learning disabilities (McKenna, Shin, & Ciullo, 2015) or gifted learners (Cassady et al., 2004). Widely-used instruments are the Classroom Assessment Scoring System (CLASS K-3) instrument (Pianta, La Paro, & Hamre, 2008), and the International Comparative Analysis of Learning and Teaching (ICALT) instrument (Van de Grift, Helms-Lorenz, & Maulana, 2014). However, only a limited amount of observation instruments focus specifically on adaptive teaching in a mainstream (secondary) classroom (Prast, Van de Weijer-Bergsma, Kroesbergen, & Van Luit, 2018).

Examples of observation instruments with this focus are the Classroom Observation Scale – Revised (COS-R) (VanTassel-Baska, Quek, & Feng, 2007), the Differentiated Classroom Observation Scale (DCOS) (Cassady et al., 2004), and the Differentiation of Mathematical Instruction (DMI) (Prast et al., 2018). Although very useful in the context of adaptive teaching, these instruments focus on specific groups or on a specific subject in primary education. The COS-R instrument is focused on gifted learners and emphasizes "the importance of concept development, thinking and reasoning, problem solving, and flexible accommodations for working with highly able learners" (VanTassel-Baska et al., 2007, p. 86). The DCOS instrument is also focused on gifted learners, but is organized differently, starting with a pre-observation interview to gain essential descriptive information, followed by the observation period and a post-observation debriefing and reflection. The observation format includes items about instructional activities, student engagement, and cognitive activities. Lastly, the DMI instrument focuses on primary education students of diverse achievement levels in mathematics and consists of whole-lesson items and fragment-specific items.

Despite the availability of these instruments, a need remains for

observation instruments that focus specifically on adaptive teaching behavior in secondary mainstream education.

### 3. Research questions

As the explicit focus on student learning in LS is believed to support teachers to become more aware of students' learning needs and adapt their teaching to these needs, this study addresses the following research questions:

- (1) To what extent do teachers participating in LS, compared to teachers who are not involved in LS, enhance their adaptive teaching practices in terms of
  - (a) teachers' intentions and perceptions toward their adaptive teaching practices?
  - (b) teachers' adaptive teaching behavior?
- (2) How do teachers' perceptions toward their adaptive teaching practices correspond or conflict with their observed adaptive teaching behavior?
- (3) How do teachers participating in LS relate their professional growth in terms of adaptive teaching behavior (if any) to the LS activities as part of the intervention?

### 4. Method

#### 4.1. Sample and procedure

The sample included in this study consists of 63 teachers from eight mainstream secondary schools in the Western and Northern part of the Netherlands.<sup>1</sup> Participants were recruited through convenience sampling as their schools were implementing LS during the data collection period. Teachers who signed up to participate in at least two LS cycles in one academic year (intervention group) as well as teachers from similar subjects in the same schools who were not involved in any LS activity (comparison group) were recruited by e-mail explaining the purpose and procedure of the research. Participation in this study for both groups was voluntary and all teachers were entitled to withdraw from the research at any time. Active informed consent was obtained from all teachers confirming they were aware of the research objectives, and that their data would be used confidentially and anonymously, and for research purposes only.

As Dutch teachers are entitled to spend 10% of their working hours on PD activities (Dutch Council for Secondary Education, 2018), the comparison group teachers could also participate in various PD activities as long as they were not involved in any LS activities. Hence, comparison group teachers were entitled to focus on adaptive teaching in their PD activities as this could also be part of the school policy, though they were not part of the LS team and their LS objectives around adaptive teaching. From the interviews with a selection of 18 of the 63 teachers (section 4.3.1), it can be derived that comparison group teachers participated in various PD activities, ranging from both formal (e.g., conferences and courses) to informal activities (e.g., discussing topics with colleagues and personal learning goals). In the context of this study it is relevant to note that one comparison group teacher participated in a four-day differentiation course (T62). This also applies to two intervention group teachers (T86, T89) who followed this differentiation course

in addition to their LS activities.

In total, 37 intervention group teachers (58.7%) and 26 comparison group teachers (41.3%) agreed to participate in this study. The sample descriptions with teachers' background characteristics are presented in Table 1. As becomes clear, the teaching context of the individual teachers in both groups could vary to a large extent in terms of their experience, qualification and teaching subject. However, independent t-tests indicated no significant differences in these baseline background characteristics assuming both groups are comparable.

The first author made planning arrangements for the classroom observations with individual teachers via e-mail or phone. Data were collected during two academic years (cohort 1: 2015–2016 and cohort 2: 2016–2017) and most teachers ( $n = 59$ ) were part of the first cohort (93.7%). The pre-test took place at the beginning of each academic year before teachers got involved in LS (September–November) and the post-test followed at the end of the academic year as soon as the LS activities officially ended, i.e., after the second LS cycle (April–June). To obtain representative data for teachers' daily practice, participants were explicitly asked to teach a 'business as usual' lesson without making any extra preparations in favor of the observed lesson. The participants and observer(s) met briefly prior to the observed lesson in order to clarify the procedure and comforting teachers not to perceive the observation as a teaching performance assessment. The actual lesson was then observed by the observer(s) and all lessons were video-recorded. Immediately after the observed lessons teachers were interviewed. In case teachers' schedules did not allow this, the interviews followed as soon as possible after the observed lessons.

Following the procedure of Van de Grift et al. (2014), the observations were carried out by observers who were trained to conduct the observations. The training for using both instruments consisted of an explanation of the instruments, group discussions, observation of three video-recorded lessons, and a thorough discussion focusing on how to rate the observation items. Candidates who met the consensus norm of 0.70 or higher on the International Comparative Analysis of Learning and Teaching (ICALT) instrument were found to be eligible to conduct observations using both instruments. All trained and selected observers received an observation manual with reference materials used in the training sessions.

All 126 lessons were observed by at least one observer. Eventually, 171 lesson observations were conducted by 16 different observers. Ten teachers (15.9%) were observed by two observers in both pre-test and post-test, 25 teachers (39.7%) were observed by three observers (i.e., two observers in pre-test or post-test), and 28 teachers (44.4%) were observed by two observers (i.e., one observer at both test moments).

#### 4.2. Lesson Study intervention

The intervention consisted of at least two LS cycles during each cohort including the essential LS features described in section 2.2. LS teams were only included in this study if their overarching LS objectives were related to addressing students' different educational needs. In these objectives there had to be at least a reference to how teachers would address students' needs through teacher instruction or teaching material. The authors assessed this criterion. Despite sharing these essential features, LS can be used in various forms suitable to different contexts (Stigler & Hiebert, 2016). Variations were found in terms of the use of 'case pupils' (Dudley, 2013), the LS team composition, time allocation, and whether an internal or external LS facilitator was consulted (Table 2). Obviously, these variations could result in different outcomes in terms of lesson plans and the way students' educational

<sup>1</sup> Students in mainstream secondary education in the Netherlands are generally placed in educational tracks based on their cognitive levels in primary education (early tracking), resulting in relatively homogeneous classrooms in terms of students' cognitive levels (OECD, 2016).

**Table 1**  
Sample descriptions ( $N = 63$ ).

	Intervention group ( $n = 37$ )	Comparison group ( $n = 26$ )
Female $n$ (%)	20 (54.1%)	13 (50.0%)
Age (in years)	$M = 44.7$ , $SD = 12.3$ (range: 21–61)	$M = 46.3$ , $SD = 13.1$ (range: 22–64)
Teaching experience (in years)	$M = 15.6$ , $SD = 11.1$ (range: 1–39)	$M = 16.9$ , $SD = 11.7$ (range: 1–37)
Teaching experience at current school (in years)	$M = 9.9$ , $SD = 7.8$ (range: 1–39)	$M = 12.3$ , $SD = 9.9$ (range: 1–32)
Teacher qualification (%)	M.Ed.: 17 (45.9%) B.Ed.: 18 (48.6%) In training: 2 (5.4%)	M.Ed.: 15 (57.7%) B.Ed.: 10 (38.5%) In training: 1 (3.8%)
Teaching subject cluster (%)	Languages: 16 (43.2%) Social sciences: 14 (37.8%) Sciences: 7 (18.9%)	Languages: 13 (50.0%) Social sciences: 11 (42.3%) Sciences: 2 (7.7%)

Notes: Teacher qualification 'in training' refers to teachers' final phase of their B.Ed./M.Ed. Teacher training program. The subcategory 'languages' consists of Dutch, English, German, and French. The subcategory 'social sciences' consists of Economics, History, and Geography. The subcategory 'sciences' consists of Mathematics, Physics, Chemistry, and Biology.

**Table 2**  
School variations in the intervention group.

School	Intervention group ( $n = 37$ )		Use of 'case pupils'	LS team composition	Time allocation in hours	LS Facilitator
	Female	Male				
1	2	2	Yes	Content specific	27	External
2	0	4	Yes	Content specific	27	External
3	1	2	Yes	Content specific	27	External
4	6	0	Yes	Content specific	27	External
5	3	5	Yes	Content specific	27	External
6	4	0	No	Interdisciplinary	166	Internal
7	2	2	No	Interdisciplinary	166	Internal
8	2	2	Yes	Interdisciplinary	85	External & internal

needs were addressed.

Six of the eight schools followed the Dutch LS model using 'case pupils' (De Vries, Verhoef, & Goei, 2016) which is based on Dudley's (2013) model used in the United Kingdom. In two LS cycles per academic year, three case pupils were selected who represent different attainment groupings based on the lesson objectives of the LS team. These lesson objectives could be focused on the subject matter or on more generic elements such student motivation. Lesson plans, 'live' research lesson observations and post lesson discussions were then organized around these 'case pupils', examining their learning needs and how to address these needs accordingly. After each research lesson, these 'case pupils' were briefly interviewed following the same procedure as used in the UK (Warwick et al., 2019). In the remaining two schools where the concept of 'case pupils' was not applied, a less procedural format was used to define students' educational needs.

In terms of the team composition, five LS teams were organized in content specific LS teams whereas three LS teams with teachers from various subject areas had a multidisciplinary focus. Variation was also found in terms of allocated time to participate in LS. Due to funding, two schools (#6 and #7) were able to schedule a fixed weekly afternoon for teachers to conduct LS activities. In the remaining schools, time allocation was partly funded from collective and personal teacher funds as part of the Dutch Collective Labor Agreement for Secondary Education. Schools #1 to #5 organized six meetings of approximately 2 h each per LS cycle, supplemented with an introduction meeting at the start and an overall reflection meeting at the end of two LS cycles. School #8 organized five meetings of 4 h per cycle, supplemented with extra time to spend on preparing the research lesson.

A last variation was found in the type of LS facilitator. Most

teams were coached by an external and trained LS facilitator from a partner university. Two schools deployed internal LS facilitators only who can be defined as teachers from the same school with experience in facilitating team processes, but not always having extensive knowledge of and experience with LS. One school used a combination of both external and internal LS facilitators.

#### 4.3. Instruments and data-analysis

Data for this study were collected through various research instruments which are explained in this section. Both self-report measures as well as observation instruments were used as a form of triangulation to increase confidence in the results (Schoenfeld, 2013). Table 3 presents an overview of the instruments used in this study, their focus and moment of data collection.

##### 4.3.1. Measuring teachers' intentions and perceptions through SRI's

As adaptive teaching entails both planned and spontaneous behavior, and requires knowledge about students, pedagogical content knowledge, skills, and an adaptive mindset (Van Geel et al., 2019), we conducted Stimulated Recall Interviews (SRI) in which participants verbalized their thoughts and actions concerning a specific teaching situation (Beers, Boshuizen, Kirschner, Gijsselaers, & Westendorp, 2008; Vesterinen, Toom, & Patrikainen, 2010). This instrument can be used for clarifying decision-making processes (Van Geel et al., 2019).

The SRI's contained 20 questions focused on teachers' intentions prior to the observed lesson (e.g., teachers' lesson objectives) as well as their thoughts and actions during the lesson, in particular about adaptive teaching (e.g., "To what extent were you able to cope with students' different educational needs in this lesson?"). These

**Table 3**  
Overview of the instruments used in this study.

Instrument	Focus instrument	Type of data	Moment of data collection
1. Stimulated Recall Interviews (SRI)	Teachers' intentions and perceptions.	Transcripts of 18 selected teachers.	Following post-test lesson observation (after LS intervention)
2. ICALT observation instrument	Six domains of teaching behavior (32 items) and one domain of student involvement (3 items).	Observation protocol.	Pre-test and post-test 'business-as-usual lesson'.
3. NRR observation instrument	Adaptive teaching behavior in two categories (23 items).	Observation protocol	Pre-test and post-test 'business-as-usual lesson' <sup>a</sup> .

Note:

<sup>a</sup> The moment of data collection with the ICALT and NRR instruments is identical. The pre-test and post-test business-as-usual lessons were observed with both these instruments during each lesson.

questions were supplemented with several questions related to teachers' PD activities during the academic year and to what extent teachers gained new knowledge and skills from these activities (e.g., "Did you show particular teaching behavior in this lesson that you have worked on during this academic year?").

The interviews were scheduled immediately after the observed post-test lesson or as soon as the teacher was available in case teachers' schedules did not allow this. The interviews lasted on average 13.32 min ( $M = 14.23$  min in the intervention group and  $M = 12.41$  min in the comparison group), and were conducted by the observer(s) who also observed the lesson using the observation instruments. Consequently, the observers could easily relate to the specific teaching situations they had witnessed.

In order to give a representative image of the entire sample, we selected the post-test interview data of 18 teachers who showed a relatively high increase, high decrease or remained relatively constant on their post-test measures as opposed to their pre-test scores. Selection was based on the difference in weighted average between pre-test and post-test. Three teachers from each category in both the intervention and comparison group were eventually selected (Table 4).

The analysis of the interviews consisted of the following steps: first the interviews were transcribed verbatim by the first author. All comments related to adaptive teaching, whether positive, negative or neutral were subsequently highlighted and categorized. The highlighted data were then related to teachers' PD activities. Comments about LS in the intervention group were highlighted as well. Finally, for each NRR category a possible explanation was sought in the data for teachers' relative increase, decrease or

constant outcome.

#### 4.3.2. Measuring teaching behavior using the ICALT instrument

The ICALT observation instrument was used to measure teaching quality as a one-dimensional linear construct and consists of six cumulatively ordered domains of teacher behavior (Van der Lans et al. 2018), ranging from lower order teaching behavior (i.e., 'creating a safe and stimulating learning climate') to higher order teaching behavior (i.e., 'adjusting the instruction and learner processing to inter-learner differences', in short: 'differentiation'). A seventh domain is focused on student involvement.

The internal consistency values were calculated for all ICALT domains based on the means of both measurement moments and the data from both observers. Additionally, Intraclass Correlation Coefficients (ICC) were calculated to assess the consistency

**Table 5**  
ICALT domains with number of items, internal consistency- and ICC-values.

ICALT domain	# items	Cronbach's Alpha ( $\alpha$ )	ICC
1. Safe and stimulating learning climate	4	.88	.70
2. Efficient classroom management	4	.84	.77
3. Clarity of instruction	7	.86	.66
4. Activating learning	7	.79	.79
5. Adaptive teaching	4	.79	.87
6. Teaching learning strategies	6	.79	.65
7. <i>Student involvement</i>	3	.91	.71

Notes: Domain 'Student involvement' is presented in italics since this domain does not explicitly focus on teacher behavior. Cronbach's alpha values were calculated as the mean of both observers in the pre-test. ICC was calculated over 32 valid pre-test observations and was conducted using a two-way random model with absolute agreement.

**Table 4**  
Selected teachers for post-test interviews.

Teacher ID	School ID	Intervention/comparison group	Category	Increase/decrease NRR	Weighted average pre-test
31	7	Intervention group	Rel. high increase	3.80	0.20
13	4	Intervention group	Rel. high increase	2.00	0.83
32	2	Intervention group	Rel. high increase	1.47	1.53
62	7	Comparison group	Rel. high increase	2.23	1.77
37	1	Comparison group	Rel. high increase	2.05	1.20
7	2	Comparison group	Rel. high increase	1.82	0.60
55	3	Intervention group	Rel. high decrease	-3.80	6.00
89	7	Intervention group	Rel. high decrease	-2.17	4.67
34	7	Intervention group	Rel. high decrease	-1.58	2.38
54	3	Comparison group	Rel. high decrease	-2.17	3.50
63	7	Comparison group	Rel. high decrease	-1.25	1.75
36	1	Comparison group	Rel. high decrease	-1.08	2.00
86	7	Intervention group	Relatively constant	0.25	1.50
2	1	Intervention group	Relatively constant	-0.43	2.31
24	6	Intervention group	Relatively constant	-0.16	2.13
58	2	Comparison group	Relatively constant	0.14	0.86
22	6	Comparison group	Relatively constant	0.10	2.02
68	1	Comparison group	Relatively constant	-0.09	1.80

Notes: Increase/decrease NRR refers to the difference in weighted average between pre-test and post-test. For each category in both groups, we selected the highest increasing/decreasing scores as well as the relatively most constant values.

between the raters. The results can be found in Table 5 and indicate (highly) reliable domains and moderate to highly reliable ICC values.

Examining the baseline differences between the intervention and comparison groups revealed a significant difference in the domain adaptive teaching behavior ( $t(61) = -2.52, p < .05$ ), showing that the mean value of the intervention group ( $M = 1.81$ ) is already higher than the pre-test mean value of the comparison group ( $M = 1.44$ ) at the start of the intervention. This is further illustrated in Table 11 of the results section 5.2.

Intervention effects were examined using repeated measures ANOVA with time (pre-test and post-test) as a within-subject variable and group (intervention group and comparison group) as a between-subjects factor. The analysis of between-group effects was succeeded by within-group analysis using paired-sampled t-tests and, due to violations of normality in several domains, their non-parametrical counterpart (Wilcoxon signed-ranks tests).

#### 4.3.3. Measuring adaptive teaching behavior using the NRR instrument

A Narrative Running Record (NRR) observation instrument was developed to measure adaptive teaching behavior in more depth. NRR's are generally used to record classroom activities and interactions during a certain period of time (e.g., a lesson), often divided into different time or event intervals (Stronge, Ward, Tucker, & Hindman, 2008). This instrument can focus on what teachers say or do, and how they respond to students' behavior, but can also contain elements such as body language and movement in the classroom (Smith, Baker, Hattie, & Bond, 2008). Although NRR instruments are characterized by its narratives, sometimes these narrative recordings "are quantified according to component ratings, or scored on the basis of judgements such as sufficient or insufficient" (Van de Grift, 2007, p. 137).

The analyses of the developed NRR instrument concentrated on a subset of adaptive teaching items. Two main categories were distinguished: (1) how teachers differentiated their instruction in the process of interacting with students, and (2) how teachers differentiated the content for different cognitive levels and students' learning preferences. Each category was subsequently divided into eight subcategories resulting in 28 items in total (Table 6). All these items were rated on a dichotomous scale: whether they were visible for the observers or not. Contrary to other NRR instruments that generally use fixed time intervals (Stronge, Ward, Tucker, & Hindman, 2008), we used intervals based on the duration of a particular lesson phase (e.g., "the teacher activates prior knowledge" or "the teacher organizes guided practice"). This was a result of pre-testing the instrument where we found that more information could be retrieved using intervals based on the duration of a particular lesson phase as opposed to using fixed time intervals that often yielded information from overlapping lesson phases.

The NRR instrument was constructed for the purpose of this study and was validated following psychometric testing. The ICALT observation instrument was used as an anchor to assess the validity of the NRR. To test the hypothesis that both the NRR and the ICALT data together fit the one-dimensional, cumulative pattern, we first assessed whether the NRR ratings showed a similar item response pattern as the ICALT. The psychometric tests indicated an overall good fit. Five items (N3, A3, F1, O1, and W3) were identified as 'misfitting' the predicted response pattern. The remaining 23 adaptive teaching items were concurrently calibrated on the same scale as the ICALT items.

The procedure for examining the data and measuring the intervention effects was consistent with the procedure of the ICALT instrument. First, we assessed the internal consistency and ICC

values (Table 7), resulting in reliable outcomes. We then examined baseline differences and subsequently tested for intervention effects. To analyze the NRR data, we used the total sum of the adaptive teaching items and calculated the weighted average using the amount of NRR items divided by the amount of time frames as indicated by the observers. We did not assess the four sub-categories within each of the two main categories due to the limited amount of items in these subcategories (resulting in lower reliability values).

Baseline differences in terms of the variables described above, resulted in one significant difference between both groups (calculated over the included 23 NRR items only). This applies to the weighted mean of the category 'Differentiating in content' ( $t(42.78) = -3.14, p < .01$ ), where the intervention group ( $M = 0.53$ ) was significantly higher than the comparison group ( $M = 0.16$ ). This is further illustrated in Table 11 of the results section 5.2.

## 5. Results

We aim to answer the research questions in four separate sections. Section 5.1 addresses teachers' intentions and perceptions toward their adaptive teaching practices (research question 1a). Section 5.2 presents the results from the observation instruments (research question 1b). In section 5.3, teachers' perceptions are related to their observed behavior (research question 2). Lastly, section 5.4 deals with the intervention group teachers' perceptions in relation to their experiences with LS (research question 3).

### 5.1. Teachers' intentions and perceptions toward their adaptive teaching practices

Teachers' intentions for the observed lessons were captured by examining the lesson objectives teachers stated in the SRI's. These lesson objectives could be clustered into content specific objectives (e.g., "that they know what coastal erosion is"), process related objectives (e.g., "making sure that everyone finished their assignment"), objectives related to learning strategies (e.g., "paying attention to self-responsible learning"), and adaptive teaching objectives (e.g., "differentiation in pace") (Table 8). Four teachers related to two categories and only one comparison group teacher referred to adaptive teaching objectives (T7). A clear difference between both groups was found in terms of process related objectives which were part of five intervention group teachers' answers as opposed to one comparison group teacher, but it is not clear whether this was influenced by LS.

The question 'what are these objectives based on?' was answered by the majority of teachers ( $n = 14$ ) in terms of (national) standards, i.e., teachers formulated their lesson objectives based on what the textbook prescribes. Two teachers, one in each group (T13 and T37), explicitly expressed that addressing students' educational needs was an element which they took into account while formulating their lesson objectives related to learning strategies.

In terms of teachers' general perceptions regarding the observed post-lesson (Table 9), seven of the interviewed teachers reported behavior that can be clustered as lower-order teaching behavior, focusing on (aspects of) classroom management and instruction (Van der Lans et al. 2018). These teachers mentioned they worked on delivering a concise instruction (T37), organizing group work for students (T2), or experimenting with new work formats (T68). Five teachers reported behavior that could be clustered as higher-order teaching behavior (Van der Lans et al. 2018) relating to how they could trigger student learning through activating and motivating students. Furthermore, two teachers referred to implementing ICT such as working with a new digital method (T55) or allowing students to use laptops during the lesson (T7), and one comparison

**Table 6**  
Extract of adaptive teaching items as part of the NRR observation instrument.

Adaptive teaching behavior	
Teacher – student interaction	Differentiating in content
<i>The teacher ...</i>	<i>The teacher ...</i>
<b>Cognitive level</b>	<b>Instruction</b>
... adapts instruction to individual student(s) by ...	- [U1] ... explains the concept or topic at different cognitive levels.
- [N1] ... providing tips and support.	- [U2] ... shows how something works in different ways and at different cognitive levels.
- [N2] ... providing support in case of experienced difficulties.	
- [N3]* ... checking the cognitive level at which they perform.	
<b>Encouragement</b>	<b>Exercises</b>
... encourages individual student(s) by ...	... differentiates the exercises during the lesson ...
- [A1] ... setting challenging goals.	- [O1]* ... for different cognitive levels.
- [A2] ... setting a clear goal together with the student.	- [O2] ... in order to be challenging for different cognitive levels.
- [A3]* ... making them think about the subject content.	- [O3] ... with respect to the workload.
- [A4] ... asking them for 'counter-examples'.	- [O4] ... with respect to time (letting some students work longer than others).
- [A5] ... asking them to consult (an)other student(s).	- [O5] ... by providing (supportive) material for some students.
- [A6] ... letting them help (an)other student(s).	
<b>Feedback</b>	<b>Work formats</b>
... gives individual students feedback by ...	... differentiates work formats ...
- [F1]* ... allowing them to respond to the subject content.	- [W1] ... using ICT resources.
- [F2] ... responding affirmatively to positive input.	- [W2] ... using formative assessment.
- [F3] ... responding to students' answer(s) with more follow-up questions.	- [W3]* ... using audiovisual material.
- [F4] ... responding to student's learning need(s) that (s)he presents.	- [W4] ... by having students of different cognitive levels work together.
<b>Learning strategies</b>	<b>Homework</b>
- [L1] ... recommends different learning strategies to students.	- [H1] ... gives students different homework exercises.
- [L2] ... let students set their own goals for the lesson.	- [H2] ... assists individual students with the planning of their homework.

Note: Five items in this extract, displayed with an asterisk (\*), were eventually removed from this extract due to misfitting patterns in the validation process. These items were not included in the data analysis.

**Table 7**  
NRR categories with number of items, internal consistency- and ICC-values.

NRR category	# items	Cronbach's Alpha ( $\alpha$ )	ICC
Adaptive teaching overall	28	.85	.85
Teacher – student interaction	15	.74	.74
Differentiating in content	13	.85	.82

Notes: Cronbach's alpha and ICC values were calculated over the sum score of NRR pre-test items. ICC pre-test values were calculated over 35 valid observations and the original amount of items ( $N = 28$ ) using a two-way mixed model with absolute agreement.

**Table 8**  
SRI categories in terms of lesson objectives.

SRI categories	Intervention group ( $n = 9$ )	Comparison group ( $n = 9$ )
1. Content specific objectives	T2, T13*, T31*, T32*, T55, T89	T7*, T22, T36, T54, T62, T63, T68
2. Process related objectives	T24, T31*, T32*, T34, T86	T58
3. Objectives related to learning strategies	T13*	T37
4. Adaptive teaching objectives		T7*

Note: Teachers denoted with an asterisk are presented in two categories.

**Table 9**  
SRI categories in terms of general teacher perceptions.

SRI categories	Intervention group ( $n = 9$ )	Comparison group ( $n = 9$ )
1. Classroom management, structure, instruction, experimenting with work formats	T34, T2, T86	T37, T36, T63, T68
2. Activating/motivating students,	T13, T32, T24	T54, T22
3. Differentiating and formative assessing	T31, T89	T62
4. Implementing ICT	T55	T7
5. No particular behavior		T58

group teacher argued that she did not show any particular teaching behavior she had been working on during the intervention period. No clear differences were found between both groups.

More specifically, in terms of teachers' perceptions toward their adaptive teaching practices, the question 'To what extent were you

able to cope with students' different educational needs in this lesson?' was answered affirmatively by thirteen teachers. These teachers reported various forms of how they adapted their teaching in the observed lessons. The answers were clustered in different categories which are presented in Table 10. Half of the interviewed teachers' answers (five intervention group teachers and four comparison group teachers) related to addressing students' individual needs. In most of these cases, teachers referred to differentiating between students' cognitive levels in pace and complexity of the instruction. Various examples were given: from allowing a high

achieving student to take full responsibility for his own learning: "He writes his own plan, let's say. And he is often working on physics [other subject] as well, which I don't mind at all" (T31), to a more teacher-led form of adaptive teaching: "Yes, I mean, those ladies [students] over there are performing very well, so I give them

**Table 10**  
SRI categories in terms of perceived adaptive teaching in observed lesson.

SRI categories	Intervention group (n = 9)	Comparison group (n = 9)
1. No adaptive teaching	T34	T54
2. Limited adaptive teaching	–	T22, T36, T58
3. Adaptive teaching at group level	T2	T7, T37*, T62*
4. Adaptive teaching at individual level	T24, T31, T32, T55, T89	T37*, T62*, T63, T68
5. Adaptive teaching in content	T13, T86	–

Note: Teachers denoted with an asterisk are presented in two categories.

**Table 11**  
Pre-test and post-test results for both the ICALT and NRR instruments.

Domain/category	Pre-test		Post-test	
	M	SD	M	SD
<i>Intervention group (n = 37)</i>				
1. Safe and stimulating learning climate	3.47	.47	3.53	.48
2. Efficient classroom management	3.11	.49	3.29*	.58
3. Clarity of instruction	2.94	.52	3.06	.57
4. Activating learning	2.50	.53	2.63	.58
5. Adaptive teaching	1.81	.68	1.98	.73
- NRR adaptive teaching overall	2.03	1.34	1.83	1.20
- NRR teacher – student interaction	1.49	.88	1.27	.70
- NRR differentiating in content	.53	.68	.56	.69
6. Teaching learning strategies	1.93	.55	2.00	.49
7. Student involvement	2.76	.72	2.81	.77
<i>Comparison group (n = 26)</i>				
1. Safe and stimulating learning climate	3.27	.59	3.43	.52
2. Efficient classroom management	3.20	.66	3.21	.61
3. Clarity of instruction	2.93	.51	2.95	.65
4. Activating learning	2.38	.55	2.57	.58
5. Adaptive teaching	1.44	.41	1.75**	.53
- NRR adaptive teaching overall	1.53	.69	1.60	.81
- NRR teacher – student interaction	1.37	.67	1.25	.47
- NRR differentiating in content	.16	.18	.35	.52
6. Teaching learning strategies	1.69	.46	2.05*	.74
7. Student involvement	2.83	.61	2.71	.72

Notes: Given the linear additive one-dimensionality pattern, we merged the data from the ICALT and the NRR in this table. The NRR data are presented under the fifth ICALT domain (adaptive teaching) and refer to the weighted means of (the sum score of) each category. \* $p < .05$  (two-tailed). \*\* $p < .01$  (two-tailed).

other tasks. I also expect more from them” (T32).

Another teacher (T68) argued that adaptive teaching is basically about involving all students in her lesson so that they are working on what they should be working on without necessarily differentiating the content between students. Other answers in this category contain elements of addressing students’ learning preferences: “There was something to listen to, there was something to see, there was something to do, there was attention for the whole group as well as more individual attention in that last part” (T63), and: “So kids who are more visually oriented are also addressed, as well as auditory oriented students. Pace, level, that’s where the differentiation is” (T55).

The cluster adaptive teaching at the group level includes four teachers (of which three comparison group teachers). An illustrative answer in this category was given by the only intervention group teacher:

“In one group they are just very ahead and then you know, OK, you can let them work quietly. But with a different group, you really have to sit down with, OK, what are the steps, what are you going to do?” (T2).

The answers of two intervention group teachers were assigned to the category adaptive teaching in content. Teacher 86 argued that she did not really adapt her teaching to individual students, yet she organized a lot of variety in the lesson tasks to make sure that

all students were involved. Teacher 13 explained that she made an assignment for students to prepare parts of the lesson in order for them to experience freedom and come up with their own ideas.

In contrast to the teachers who responded affirmatively to the question to what extent they were able to cope with students’ different educational needs, five teachers, one of whom was an intervention group teacher, and four comparison group teachers answered this question negatively. Two of these teachers firmly said “no” (T54, comparison group) or “no, not in this lesson” (T34, intervention group). The three other comparison group teachers (T22, T36, and T58) argued that they only paid limited attention to differences between students. Teacher 36 explained this in terms of allowing a student to work on extra questions: “Well actually, only that I said that if you [student] are finished, you do the two extra questions, because those were the more difficult questions”. Teacher 58 referred to supporting a student with extra instruction who was ill during the previous lesson and teacher 22 argued that the observed lesson was “obviously a little differentiated”, however, he continues that “the last two, three lessons, I paid a lot of individual attention to students, so [I] really looked at what does that student need ...” and later “I’m not an advocate of differentiating all the time and in every lesson”.

## 5.2. Teachers’ (adaptive) teaching behavior

This section presents the observation results from both pre-test and post-test measurements. Table 11 provides the descriptive statistics derived from the ICALT and NRR observation instruments. The results are intentionally merged into one table to illustrate its additive one-dimensionality pattern. Both groups showed increases on all ICALT domains and on one of the NRR categories. The intervention group reported increasing post-test results on all ICALT domains, of which one is a significant increase (i.e., ‘efficient classroom management’). However, this group also showed two decreasing NRR categories (not significant). Contrary to the assumption that a comparison group in (quasi-) experimental research acts as a baseline (Field, 2013), the comparison group in this study reports significant increases on two ICALT domains (i.e., ‘adaptive teaching’ and ‘teacher learning strategies’). It is noteworthy that the adaptive teaching pre-test value of the intervention group is already significantly higher than the pre-test value of the comparison group. Moreover, this pre-test value is even higher than the increased post-test value of the comparison group for this domain.

Repeated-measures ANOVA tests were conducted to examine the differences between groups during the intervention period. This yielded no significant differences implying that based on these observation results, we cannot argue that LS leads to significantly more observed adaptive teaching behavior. The SRI could potentially elicit underlying motives for teachers’ increase, decrease or stable situation on the post-test compared to their pre-test values.

### 5.3. Relating teachers' answers to their NRR category

Following the main findings of teachers' self-report and observation data, in this section we examine whether the NRR output, in terms of teachers' growth (or decline) on the NRR in the post-test (Table 4 in section 4.3.1), corresponds or conflicts with teachers' perceptions about their adaptive teaching. The five teachers who argued that their observed post-test lesson did not contain any, or only limited, adaptive teaching teacher behavior (Table 10), were found in the decreasing or constant NRR categories: 'relatively high decrease' (T34, T54, T36) or 'relatively constant' (T22, T58). As Table 10 shows, only one of these teachers is part of the intervention group. Examining these teachers' post-test weighted NRR averages, teacher 22 draws specific attention given a relatively high score on observed adaptive teaching (though relatively constant over time), but he was placed in the 'limited adaptive teaching category' as a result of his expressed perceptions about his adaptive teaching behavior.

Teachers who were found in the 'relatively high increase' NRR category argued that their lessons contained elements of adaptive teaching behavior. In the intervention group, all three teachers gave clear examples of how they adapted their teaching to the individual needs of students. In particular, teachers 13 and 31 showed clear signs of becoming more aware of students' differences and how they addressed these differences. Teacher 31 argued that "I've become more aware of the individual differences", but commented that it was perhaps not very visible in the post-lesson although he explained in detail how he dealt with different students and how these students could be characterized in terms of capabilities, results and other educational needs. Teacher 13 stated that adaptive teaching is something she is always focused on, also prior to the intervention period:

"No, I believe this is actually something I always do, yes. [...] I have students in my classroom, that's something you were just able to see, who get their own worksheets from me. They can often apply their own pace as well. [...] And so they [students] can indicate themselves at which pace do I work, do I immerse myself?" (T13).

Conflicting situations between the NRR categories and teachers' perceptions were also found, especially in the category 'relatively high decrease'. Three teachers, two of whom were intervention group teachers (T55 and T89) and one comparison group teacher (T63), argued that their observed post-lesson contained (plenty of) adaptive teaching behavior. Teacher 55, for example, elaborated on the use of a new digital method which he referred to as an 'adaptive method'. He explained how this method yields a lot of student information in terms of formative test scores and how he can subsequently guide and support students. Moreover, he argued that:

"Since this new digital method is in use, I teach very differently than I used to in the past. In the past, I was giving 40 minutes of instruction, explaining, and that's currently not the case at all. I have not given any instruction now [in the observed post-lesson]. I've only been guiding [students]".

For teacher 89 a similar situation applies. Her interview contained ample references to adaptive teaching in the LS process and how this influenced her teaching: "We have been working a lot on this as well in LS, in the two research lessons. It went indeed about responding more to [students'] needs. This is another way. I'm working on this more and more". The interview of comparison group teacher 63 contains remarks about adaptive teaching, but not

as convincingly as the two intervention group teachers (as illustrated by the quote in section 5.1).

### 5.4. Relating teachers' answers to their LS participation

To examine whether and to what extent teachers relate their answers to their participation in LS, we draw on the SRI data of the intervention group. All these nine teachers reported mostly positive experiences with LS. These main points can be clustered in four categories (Table 12): (1) learning from each other through discussion, exchanging information, and observation; (2) experimenting with new teaching elements; (3) focusing on adaptive teaching; and (4) gaining a critical stance, learning about subject matter, and changes in teacher behavior. These categories can be interconnected as teacher 31 illustrated: "New knowledge, no. Insights, yes, that I actually teach the same way for the last 30 years and that a few things could surely have been done differently", which, according to him, is enabled "through the conversations with others, my colleagues, in LS. I actually liked that part most. Discussing lessons like how do you do this and how do you do that" [...] "And experimenting, just conducting experiments".

The first category was referred to by five teachers. They appreciated working closely together to learn from each other and to get to know each other (better). In-depth discussion and classroom observation were often cited as the most important elements of the collaborative work. The strength of live classroom observation was also addressed by various teachers. Similarly, experimenting in LS (the second category) was valued by two teachers. Teacher 13 illustrated that she learned in particular from the failed experiments: "Well, I think it's nice to see that some things are not working then. I can really find this useful".

The adaptive teaching category contains three teachers. Teacher 13 explained how the focus on differentiation in their LS team sparked their enthusiasm. Teacher 24 cited working with 'case pupils' as an eye-opener, but does not really elaborate on this. Teacher 89, however, is very specific:

"We actually came to the conclusion that if you become aware of the fact that there are students' needs, that you indeed have to focus on one or two learning needs per students, and especially, but I already knew this about differentiation, is that differentiation is something you always do prior to the lesson".

Six teachers were clustered in the last category about various elements of professional growth. Teacher 13 reported that participating in LS made her realize that "as a teacher, you can really want your students to become owners [of the content], but you have to place it more with the students themselves. And that's what I try to do more". Teachers 34 and 55 reported that LS helped them to prepare students differently for the national central exams. In the case of teacher 34 the way she helped students in their reasoning was picked up by a teacher from another school who complimented her for the excellent way students approached the questions.

Teacher 32 argued that, as a result of LS, "you are just working differently on your lessons. You're not that rigid anymore. If I would make a study planner for next year now, I would do it differently than I would have always done it" and she also reported: "you become critical towards your own teaching". A similar notion was given by teacher 89, but this teacher referred more broadly to the effects (at meso level), arguing that LS came "exactly at the right moment because we, as a school, are in a transformation as well".

In contrast, two teachers also expressed their negative experiences with LS. Teacher 13 argued that their LS facilitator was not treating the LS team as professionals and teacher 55 disliked the fact that the research lessons were more focused on a traditional,

**Table 12**  
Positive experiences of participating in Lesson Study.

Teacher ID	Increase/decrease NRR (baseline)	Positive experiences with LS [cluster]
T31	3.80 (.20)	- Learning from colleagues through discussion and exchanging information [1] and experimenting [2] - Insights about his own teaching [4]
T13	2.00 (.83)	- Working collaboratively with colleagues: discussion with colleagues in particular [1] - Experimenting with new elements even when it fails [2] - Focus on differentiated teaching is really helpful [3] - Professional growth in terms of insights and behavior [4]
T32	1.47 (1.53)	- Critical stance towards your own teaching and working differently as a result of LS [4]
T86	0.25 (1.50)	- Getting closer to colleagues and classroom observation [1] - Learned a lot as a team [4]
T2	-.06 (2.31)	- Discussion with and learning from each other [1]
T24	-.16 (2.13)	- Working with case students [3]
T34	-1.58 (2.38)	- To get to know your colleagues better (as a starting teacher) [1] - Learned about subject content (reasoning). [4]
T89	-2.17 (4.67)	- Very positive about LS in general and its focus on adaptive teaching in particular [3] - Really put something in motion in school [4]
T55	-3.80 (6.00)	- Learned about subject content [4]

Note: Baseline value refers to teachers' NRR weighted average pre-test score. The output is arranged by teachers' NRR weighted average increase/decrease value in the post-test as opposed to pre-test, starting with the relatively highest NRR increase and ending with the relatively highest decrease.

frontal way of teaching whereas his lessons were more organized around a digital method where this way of teaching is not suitable.

## 6. Conclusion and discussion

The aim of this study was to examine whether participating in the increasingly popular PD approach LS influences teachers' adaptive teaching practices in favor of students' different learning needs. A relatively unique research design in the context of LS research was deployed to measure teachers' perceptions about their adaptive teaching practices as well as their adaptive teaching behavior.

Overall, the results revealed that LS played a pivotal role in influencing teachers' perceptions, showing how teachers have become more aware of students' different educational needs and how they address these needs accordingly. This corroborates earlier findings (Schipper et al. 2017; Schipper et al. 2018; Norwich & Ylonen, 2013). In contrast, no significant intervention effects were found in the observation data suggesting that intervention group teachers do not behave differently in the post-test lessons. Answering the three research questions may clarify these seemingly contradictory outcomes.

The first research question, to what extent participating in LS leads to increased adaptive teaching behavior, was answered in two ways. Firstly, teachers' intentions and perceptions about their adaptive teaching practices were examined using the post-lesson interview data. The results showed that teachers do not seem to refer to adaptive teaching in their lesson objectives (with the exception of one comparison group teacher). This is arguably related to the way teachers in Dutch secondary education are used to formulate lesson objectives, often focusing on content specific objectives. However, teachers' perceptions of their post-test lessons revealed ample evidence of how they addressed students' educational needs. Intervention group teachers tended to give more notion of their adaptive teaching practices by expressing clear examples of their post-test lesson behavior. Nevertheless, teachers from both groups defined and described adaptive teaching in a variety of ways (e.g., in terms of motivation, ability, learning preferences) and at different levels (e.g., individuals or groups). This is consistent with the literature showing that a clear and delineated definition of adaptive teaching is lacking (e.g., Parsons et al., 2018; Suprayogi et al., 2017).

Secondly, as stated in the second paragraph of this section, no intervention effects were found in terms of observed teacher behavior. Despite this, several within-group differences in both groups were reported. Intervention group teachers presented

significantly higher levels of efficient classroom management. In the comparison group, on the other hand, significantly higher outcomes were found in terms of adaptive teaching and teaching learning strategies. The significant increases in the comparison group are remarkable and it is unclear what caused this. What does become clear, however, is that the average adaptive teaching pre-test value of this group was already significantly lower than the average pre-test score of the intervention group. Moreover, even the average post-test score of the comparison group is still lower than the average pre-test score of the intervention group. A possible reason could be related to the formal or informal PD activities conducted by the comparison group teachers. However, given the large variety of PD activities in this group, this does not seem obvious. Another possible reason could relate to the limited sample size or to the statistical phenomenon 'regression to the mean' (Barnett, Van der Pols, & Dobson, 2005) given the relatively low pre-test value.

The second research question related teachers' perceptions to their growth (or decline) in terms of their NRR post-test output. Teachers who argued that their post-test lesson contained no or limited elements of adaptive teaching were also found in the relatively high decreasing or constant NRR categories, which suggests that these findings correspond. This also applies to teachers in the relatively high increasing NRR category who reported with clear examples of how they gained awareness of students' educational needs and how they addressed these needs accordingly. In other cases, conflicting situations arose between teachers' perceptions and their post-test NRR output, especially teachers in the decreasing NRR category who reported ample examples of their adaptive teaching. Clearly, their decline in observation outcomes over the course of the intervention period did not match their perceptions in the post-test.

The third and last research question related the perceptions of intervention group teachers to their participation in LS. The results showed that this subsample of nine teachers was predominantly positive about LS and reported how they learned from their LS team members and how they experimented with new teaching elements. Teachers also argued that LS supported them in focusing on students' educational needs. The concept of 'case pupils' was highly valued in particular and perceived as an eye-opener. As a result of working with LS they reported structural changes in terms of a critical stance and (adaptive) teaching behavior.

In addition to these results, this study also reports on the validation of an observation instrument focusing specifically on adaptive teaching. The psychometric analyses revealed that the NRR instrument indeed measures adaptive teaching and that it can

therefore be used in other (research) contexts as well.

The strength of this study is that it was situated in different educational contexts (eight schools) and that it approaches the effects of participating in LS from both a teacher perspective as well as an observer perspective, yielding a 'fresh picture' of teachers' adaptive teaching practices. As this study is one of the first of its kind in terms of research design in the context of LS research, the results also raise several questions, particularly relating to how both researchers and teachers define adaptive teaching. The data showed that teachers do not always perceive their behavior as adaptive whereas the observation instruments did indicate adaptive teaching behavior, and vice versa. We also found that some teachers refer to adaptive teaching in terms of individualized instruction. However, following the concept of Corno (2008) we argue that adaptive teaching takes place in the social context of a whole-class activity by addressing students' needs in the interaction with students.

A second question is whether researchers can fully capture adaptive teaching by observing teachers in two lessons, without knowing the students and their learning capabilities, background characteristics, and previous experiences with the subject and teacher. This is consistent with the findings of Van Geel et al. (2019) who argue that "differentiation during the lesson cannot be isolated from the phases of lesson preparation and evaluation" (p. 9).

Furthermore, as pedagogical content knowledge is believed to be important in adaptive teaching (Beltramo, 2017), we wonder whether the observers' backgrounds would matter in observing adaptive teaching behavior as they were not always subject matter experts of the lessons they observed.

Although this study has a powerful design and produces meaningful results, we came across several limitations. Firstly, despite collecting evidence about teachers' LS experiences, we did not examine the actual activities and conversations in the LS meetings. As a consequence, the study was unable to control for variation in the quality with which LS was implemented. The evidence related to research question 1b should be interpreted with this limitation in mind. Also, we did not examine the school culture and conditions that supported or hindered their learning (Schipper et al. 2019). Hence, it could be that teachers who participated in LS were willing to increase their adaptive teaching behavior but felt hindered by colleagues in their LS team or the school context. Additionally, it would be worthwhile to examine whether the school climate in which LS takes place fosters inclusive education (Roy et al., 2013).

Secondly, remarkable differences were found between the observation and interview results of several teachers, and it is unclear what exactly caused this. It could be over- or underestimation of teachers (Woolfolk Hoy & Spero, 2005), but could also indicate that the observation instruments did not fully capture adaptive teaching. It could also mean that participation in the one-year LS intervention influences teachers' perceptions and awareness in the first place, which is obviously an important step, whereas changing adaptive teaching behavior would need more time.

Thirdly, despite the fact that a large amount of lessons were observed by two observers and the consensus norm was met by all observers, the majority of lessons were observed by one observer. This may have consequences for the validity.

Lastly, there may have been variation in outcomes between LS teams and schools as well as variation in (clusters of teachers') background characteristics such as teaching experience, skills, or motivation to adapt one's teaching to students' educational needs, but given the limited sample size we could not control for this. Also the variation within the intervention group in terms of facilitated time to participate in LS, differences in subject areas, use of 'case pupils', and the use of an external LS facilitator could have affected the outcomes. Consequently, we did not fully capture the variation

between these subgroups and how their contexts may have yielded different outcomes (Bryk, Gomez, Grunow, & LeMahieu, 2015). This by no means implies that LS is found ineffective, but stresses the complexity of changing (and measuring) higher order teaching skills such as adaptive teaching.

As this research is one of the first attempts to examine the effects of participating in LS on teachers' adaptive teaching practices, future research may address the questions and limitations outlined above, starting with further clarifying the "fuzzy" construct (Deunk et al., 2018, p. 32) of adaptive teaching.

## Funding

This work was supported by The Netherlands Organisation for Scientific Research (NWO) under Grant Number: 023.005.105.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.tate.2019.102961>.

## References

- Barnett, A. G., Van der Pols, J. C., & Dobson, A. J. (2005). Regression to the mean: What it is and how to deal with it. *International Journal of Epidemiology*, 34, 215–220.
- Beers, P. J., Boshuizen, H. P. A., Kirschner, P. A., Gijssels, W., & Westendorp, J. (2008). Cognitive load measurements and stimulated recall interviews for studying the effects of information and communications technology. *Educational Technology Research & Development*, 56(3), 309–328.
- Bell, C. A., Dobbelaer, M. J., Klette, K., & Visscher, A. (2019). Qualities of classroom observation systems. *School Effectiveness and School Improvement*, 30(1), 3–29.
- Beltramo, J. L. (2017). Developing adaptive teaching practices through participation in cogenerative dialogues. *Teaching and Teacher Education*, 63, 326–337.
- Bjuland, R., & Helgevd, N. (2018). Dialogic processes that enable student teachers' learning about pupil learning in mentoring conversations in a Lesson Study field practice. *Teaching and Teacher Education*, 70, 246–254.
- Bruggink, M., Goei, S. L., & Koot, H. M. (2016). Teachers' capacities to meet students' additional support needs in mainstream primary education. *Teachers and Teaching*, 22(4), 448–460.
- Brühwiler, C., & Blatchford, P. (2011). Effects of class size and adaptive teaching competency on classroom processes and academic outcome. *Learning and Instruction*, 21, 95–108.
- Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. G. (2015). *Learning to improve. How America's schools can get better at getting better* (5<sup>th</sup> ed.). Cambridge, MA: Harvard Education Press.
- Cassady, J. C., Speirs Neumeister, K. L., Adams, C. M., Cross, T. L., Dixon, F. A., & Pierce, R. L. (2004). The differentiated classroom observation scale. *Roeper Review*, 26(3), 139–146.
- Corno, L. (2008). On teaching adaptively. *Educational Psychologist*, 43(3), 161–173.
- De Vries, S., Verhoef, N., & Goei, S. L. (2016). *Lesson Study. Een praktische gids voor het onderwijs. [Lesson Study. A practical guide for education]*. Antwerp/Apeldoorn: Garant Publishers.
- Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, 38(3), 181–199.
- Deunk, M. I., Smale-Jacobse, A. E., De Boer, H., Doolaard, S., & Bosker, R. J. (2018). Effective differentiation practices: A systematic review and meta-analysis of studies on the cognitive effects of differentiation practices in primary education. *Educational Research Review*, 24, 31–54.
- Dixon, F. A., Yssel, N., McConnell, J. M., & Hardin, T. (2014). Differentiated instruction, professional development, and teacher efficacy. *Journal for the Education of the Gifted*, 37(2), 111–127.
- Dudley, P. (2013). Teacher learning in Lesson Study: What interaction-level discourse analysis revealed about how teachers utilised imagination, tacit knowledge of teaching and fresh evidence of pupils learning, to develop practice knowledge and so enhance their pupils' learning. *Teaching and Teacher Education*, 34, 107–121.
- Dudley, P. (2015). How Lesson Study works and why it creates excellent learning and teaching. In P. Dudley (Ed.), *Lesson Study. Professional learning for our time* (pp. 1–28). New York, USA: Routledge.
- Dutch Council for Secondary Education. (2018). *CAO VO 2018/2019. Collectieve arbeidsovereenkomst voor het voortgezet onderwijs. [Collective Labor Agreement for Secondary Education 2018/2019]*. Utrecht, The Netherlands: Dutch Council for Secondary Education.
- Field, A. (2013). *Discovering statistics using IBM SPSS Statistics* (4<sup>th</sup> ed.). London: Sage.
- Gehrke, R. S., & Cocchiarella, M. (2013). Preservice special and general educators' knowledge of inclusion. *Teacher Education and Special Education*, 36(3),

- 204–216.
- Godfrey, D., Seleznyov, S., Anders, J., Wollaston, N., & Barrera-Pedemonte, F. (2019). A developmental evaluation approach to lesson study: Exploring the impact of lesson study in London schools. *Professional Development in Education*, 45(2), 325–340.
- Hadfield, M., & Jopling, M. (2016). Problematising lesson study and its impacts: Studying a highly contextualised approach to professional learning. *Teaching and Teacher Education*, 60, 203–214.
- Helgevoid, N., Næsheim-Bjørkvik, G., & Østrem, S. (2015). Key focus areas and use of tools in mentoring conversations during internship in initial teacher education. *Teaching and Teacher Education*, 49, 128–137.
- Hertberg-Davis, H. (2009). Myth 7: Differentiation in the regular classroom is equivalent to gifted programs and is sufficient. Classroom teachers have the time, the skill, and the will to differentiate adequately. *Gifted Child Quarterly*, 53(4), 251–253.
- Huang, R., & Shimizu, Y. (2016). Improving teaching, developing teachers and teacher educators, and linking theory and practice through lesson study in mathematics: An international perspective. *ZDM Mathematics*, 48, 393–409.
- Jordan, A., & McGhie-Richmond, D. (2014). Identifying effective teaching practices in inclusive classrooms. International perspectives on inclusive education. In C. Forlin, & T. Loreman (Eds.), *Measuring inclusive education* (vol. 3, pp. 133–162). Bingley, UK: Emerald Group Publishing Limited.
- Kyriakides, L., Creemers, B. P. M., & Antoniou, P. (2009). Teacher behaviour and student outcomes. Suggestions for research on teacher training and professional development. *Teaching and Teacher Education*, 25, 12–23.
- Leavy, A. M., & Hourigan, M. (2016). Using lesson study to support knowledge development in initial teacher education: Insights from early number classrooms. *Teaching and Teacher Education*, 57, 161–175.
- Lee, C. K.-E. (2015). Examining education rounds through the lens of lesson study. *International Journal of Educational Research*, 73, 100–106.
- Lee Bae, C., Hayes, K. N., Seitz, J., O'Connor, D., & DiStefano, R. (2016). A coding tool for examining the substance of teacher professional learning and change with example cases from middle school science lesson study. *Teaching and Teacher Education*, 60, 164–178.
- Lewis, C., & Perry, R. (2017). Lesson study to scale up research-based knowledge: A randomized, controlled trial of fractions learning. *Journal for Research in Mathematics Education*, 48(3), 261–299.
- Lewis, C., Perry, R., & Murata, A. (2006). How should research contribute to instructional improvement? The case of lesson study. *Educational Researcher*, 35(3), 3–14.
- Maeng, J. L., & Bell, R. L. (2015). Differentiating science instruction: Secondary science teachers' practices. *International Journal of Science Education*, 37(13), 2065–2090.
- McKenna, J. W., Shin, M., & Ciullo, S. (2015). Evaluating reading and mathematics instruction for students with learning disabilities: A synthesis of observation research. *Learning Disability Quarterly*, 38(4), 195–207.
- Naraian, S., & Schlessinger, S. (2018). Becoming an inclusive educator: Agentive maneuverings in collaboratively taught classrooms. *Teaching and Teacher Education*, 71, 179–189.
- Ni Shuilleabhain, A., & Seery, A. (2018). Enacting curriculum reform through lesson study: A case study of mathematics teacher learning. *Professional Development in Education*, 44(2), 222–236.
- Norwich, B., Dudley, P., & Ylonen, A. (2014). Using lesson study to assess pupils' learning difficulties. *International Journal for Lesson and Learning Studies*, 3(2), 192–207.
- Norwich, B., Fujita, T., Adlam, A., Milton, F., & Edwards-Jones, A. (2018). Lesson study: An inter-professional collaboration approach for educational psychologists to improve teaching and learning. *Educational Psychology in Education*. <https://doi.org/10.1080/02667363.2018.1468733>.
- Norwich, B., & Ylonen, A. (2013). Design based research to develop the teaching of pupils with moderate learning difficulties (MLD): Evaluating lesson study in terms of pupil, teacher and school outcome. *Teaching and Teacher Education*, 34, 162–173.
- OECD. (2016). Netherlands 2016: Foundations for the future. *Reviews of national policies for education*. Paris: OECD Publishing. <https://doi.org/10.1787/9789264257658-en>.
- Parsons, S. A., Vaughn, M., Scales, R. Q., Gallagher, M. A., Parsons, A. W., Davis, S. G., et al. (2018). Teachers' instructional adaptations: A research synthesis. *Review of Educational Research*, 88(2), 205–242.
- Perry, R., & Lewis, C. (2009). What is successful adaptation of lesson study in the US? *Journal of Educational Change*, 10(4), 365–391.
- Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher*, 38(2), 109–119.
- Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment scoring System: Manual K-3*. Baltimore, MD: Paul H. Brookes.
- Prast, E. J., Van de Weijer-Bergsma, E., Kroesbergen, E. H., & Van Luit, J. E. H. (2018). Teaching students with diverse achievement levels: Observed implementation of differentiation in primary mathematics education. In E. J. Prast (Ed.), *Differentiation in primary mathematics education (Doctoral thesis)* (pp. 53–81). Utrecht, The Netherlands: Utrecht University. Retrieved from <https://dspace.library.uu.nl/bitstream/1874/364287/1/Prast.pdf>.
- Roy, A., Guay, F., & Valois, P. (2013). Teaching to address diverse learning needs: Development and validation of a differentiated instruction scale. *International Journal of Inclusive Education*, 17(11), 1186–1204.
- Schipper, T., Goei, S. L., De Vries, S., & Van Veen, K. (2017). Professional growth in adaptive teaching competence as a result of Lesson Study. *Teaching and Teacher Education*, 68, 289–303.
- Schipper, T., Goei, S. L., De Vries, S., & Van Veen, K. (2018). Developing teachers' self-efficacy and adaptive teaching behaviour through lesson study. *International Journal of Educational Research*, 88, 109–120.
- Schipper, T. M., De Vries, S., Goei, S. L., & Van Veen, K. (2019). Promoting a professional school culture through Lesson Study? An examination of school culture, school conditions, and teacher self-efficacy. *Professional Development in Education*. <https://doi.org/10.1080/19415257.2019.1634627>.
- Schoenfeld, A. H. (2013). Classroom observations in theory and practice. *ZDM Mathematics Education*, 45, 607–621.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14.
- Smit, R., & Humpert, W. (2012). Differentiated instruction in small schools. *Teaching and Teacher Education*, 28, 1152–1162.
- Smith, T. W., Baker, W. K., Hattie, J., & Bond, L. (2008). Chapter 12. A validity study of the certification system of the national board for professional teaching standards. In L. Ingvarson, & J. Hattie (Eds.), *Assessing teachers for professional certification: The first decade of the national board for professional teaching standards. Advances in program evaluation series (volume 11)* (pp. 345–378). Oxford: Elsevier.
- Stigler, J. W., & Hiebert, J. (2016). Lesson study, improvement, and the importing of cultural routines. *ZDM Mathematics Education*, 48, 581–587.
- Stronge, J. H., Ward, T. J., Tucker, P. D., & Hindman, J. L. (2008). What is the relationship between teacher quality and student achievement? An explorative study. *Journal of Personnel Evaluation in Education*, 20(3–4), 165–184.
- Suprayogi, M. N., Valcke, M., & Godwin, R. (2017). Teachers and their implementation of differentiated instruction in the classroom. *Teaching and Teacher Education*, 67, 291–301.
- Takahashi, A., & McDougal, T. (2016). Collaborative lesson research: Maximizing the impact of lesson study. *ZDM Mathematics Education*, 48(4), 513–526.
- Tomlinson, C. A., Brighton, C., Hertberg, H., Callahan, C. M., Moon, T. R., Brimjoin, K., et al. (2003). Differentiating instruction in response to student readiness, interest, and learning profile in academically diverse classrooms: A review of the literature. *Journal for the Education of the Gifted*, 27(2/3), 119–145.
- Tomlinson, C. A., & Imbeau, M. B. (2010). *Leading and managing a differentiated classroom*. Alexandria, Virginia, US: ASCD.
- UNESCO. (2009). *Policy guidelines on inclusion in education*. Paris, France: United Nations Educational, Scientific and Cultural Organization.
- UNESCO. (2017). *A guide for ensuring inclusion and equity in education*. Paris, France: United Nations Educational, Scientific and Cultural Organization.
- Van der Lans, Van de Grift, & Van Veen, K. (2018). Developing an instrument for teacher feedback: Using the Rasch model to explore teachers' development of effective teaching strategies and behaviors. *The Journal of Experimental Education*, 86(2), 247–264.
- Van Geel, M., Keuning, T., Frèrejean, J., Dolmans, D., Van Merriënboer, J., & Visscher, A. J. (2019). Capturing the complexity of differentiated instruction. *School Effectiveness and School Improvement*, 30(1), 51–67.
- Valli, L., & Buese, D. (2007). The changing roles of teachers in an era of high-stakes accountability. *American Educational Research Journal*, 44(3), 519–558.
- Van de Grift, W. (2007). Quality of teaching in four European countries: A review of the literature and application of an assessment instrument. *Educational Research*, 49(2), 127–152.
- Van de Grift, W., Helms-Lorenz, M., & Maulana, R. (2014). Teaching skills of pupil teachers: Calibration of an evaluation instrument and its value in predicting pupil academic engagement. *Studies In Educational Evaluation*, 43, 150–159.
- VanTassel-Baska, J., Quek, C., & Feng, A. X. (2007). The development and use of a structured teacher observation scale to assess differentiated best practice. *Roeper Review*, 29(2), 84–92.
- VanTassel-Baska, J., & Stambaugh, T. (2005). Challenges and possibilities for serving gifted learners in the regular classroom. *Theory into Practice*, 44(3), 211–217.
- Vesterinen, O., Toom, A., & Patrikainen, S. (2010). The stimulated recall method and ICTs in research on the reasoning of teachers. *International Journal of Research and Method in Education*, 33(2), 183–197.
- Vogt, F., & Rogalla, M. (2009). Developing adaptive teaching competency through coaching. *Teaching and Teacher Education*, 25, 1051–1060.
- Wake, G., Swan, M., & Foster, C. (2016). Professional learning through the collaborative design of problem-solving lessons. *Journal of Mathematics Teacher Education*, 19, 243–260.
- Wan, S. W.-Y. (2016). Differentiated instruction: Hong Kong prospective teachers' teaching efficacy and beliefs. *Teachers and Teaching*, 22(2), 148–176.
- Warwick, P., Vrikki, M., Karlsen, A. M. F., Dudley, P., & Vermunt, J. D. (2019). The role of pupil voice as a trigger for teacher learning in Lesson Study professional groups. *Cambridge Journal of Education*, 49(4), 435–455.
- Westwood, P. (2018). *Inclusive and adaptive teaching: Meeting the challenge of diversity in the classroom* (2<sup>nd</sup> ed.). London, UK: Routledge.
- Woolfolk Hoy, A., & Spero, R. B. (2005). Changes in teacher efficacy during the early years of teaching: A comparison of four measures. *Teaching and Teacher Education*, 21, 343–356.
- Xu, H., & Pedder, D. (2015). Lesson study: An international review of the research. In P. Dudley (Ed.), *Lesson Study. Professional learning for our time* (pp. 29–58). New York, USA: Routledge.