Kirsti Klette
  University of Oslo
Marte Blikstad-Balas
  University of Oslo
Astrid Roe
  University of Oslo

Linking Instruction and Student Achievement -Research design for a new generation of classroom studies

Abstract
Educational research into instructional quality would benefit from macro- and meso-level instructional data – such as achievement data or large-scale student surveys – in relation to data from the micro level – such as detailed analyses of classroom practices. Several scholars have specifically asked for studies that correlate achievement data with records of learning processes and teaching strategies, and ongoing projects attempting to do so have shown promising results. Linking different data sources on instructional quality is quite demanding because it requires a concerted effort by researchers from different fields of expertise and different traditions. A main ambition of our ongoing research project is precisely to advance such integration. As the title of the project reveals, we are dedicated to Linking Instruction and Student Achievement (LISA). In this article, we start by providing a theoretical background and status of knowledge related to instructional quality. We go on to argue that video data has shown particular promise in studies aiming to obtain systematic data from a range of classrooms in order to compare classroom practices. We then present the three components of the LISA project's design – student perception surveys, systematic classroom observation, and achievement gains in national tests – and the value of combining these three data sources. Finally, we will outline some of our findings thus far and point to future research possibilities.

Key words: instructional quality; classroom practices; video studies; mathematics; language arts

Å koble undervisning med elevprestasjoner
- Forskningsdesign for en ny generasjon klasseromsstudier

Sammendrag

For å studere undervisningskvalitet vil det være en fordel å kombinere data fra et makro og meso- nivå med detaljerte studier av hva som skjer i klasserommet. Flere har etterlyst studier som ser på sammenhenger mellom målbar faglig fremgang og lærerens undervisning. Å få til slike studier er krevende, da det forutsetter et tett samarbeid mellom forskere fra ulike felt med ulik ekspertise innenfor nokså ulike forskningstradisjoner. En hovedambisjon i vårt pågående forskningsprosjekt er nettopp å få til en slik integrasjon. Som tittelen avslører, er vi dedikert til «Linking Instruction and Student Achievement (LISA)». I denne artikkelen presenterer vi det teoretiske og empiriske grunnlaget knyttet til undervisningskvalitet. Videre argumenterer vi for verdien av videodata i studier som sammenligner undervisningspraksiser fra ulike klasserom på en systematisk måte. Deretter presenterer vi de tre datakildene i LISA-prosjektets forskningsdesign – spørreskjemaer til elever om deres oppfatninger om lærerens undervisning, systematiske klasseromsobservasjoner, og målt fremgang på nasjonale prøver i lesing og regning. Verdien av å kombinere nettopp disse tre datakildene vil også bli diskutert. Avslutningsvis deler vi noen av våre tidlige forskningsfunn.

Nøkkelord: undervisningskvalitet; klasseromspraksis; video studier; matematikk; norskfaget

Introduction: the need for links between different measures of instructional quality

Researchers agree that education is a key factor in shaping global economic and social development, and education is argued to enable peace, democracy, and equity, to eradicate poverty, and to drive sustainable development (OECD, 2010, 2016; UNESCO, 2017). In line with this, many countries have given priority to research on what happens in the classroom and how teaching and instruction influences students' learning. Recent reviews suggest that teachers' instructional practices make a difference to students' learning and that this one factor is more important than class size, classroom climate, and teachers' years of experience and formal training (Baumert et al., 2010; Bryk, Sebring, Allensworth, Easton, & Luppescu, 2010; Hattie, 2009; Konstantopoulos & Chung, 2011; Seidel & Shavelson, 2007; Timperley & Alton-Lee, 2008). As Lipowsky and colleagues (2009, p. 527) argue, recent studies indicate that the impact of both teacher characteristics and instruction is stronger than previously assumed. While there

seems to be agreement that teacher instruction matters, the field still lacks knowledge about how and why teachers' different instructional practices relate to students' learning.

The educational field has been slowed by fragmentation, a multitude of small-scale studies, and competing methodological and theoretical approaches; attempts to integrate different qualitative perspectives on instruction and quantitative measures of students' outcomes have been scarce. While international large-scale assessments like the Programme for International Student Assessment (PISA) and the Trends in International Mathematics and Science Study (TIMSS) provide quality data on student achievement, they have only to a modest degree been able to address issues of instructional quality. To gather contextual information, both PISA and TIMSS ask students to respond to background questionnaires seeking information about classroom and school climate and aspects of learning and instruction, including students' interest, motivation, and engagement (OECD, 2016; Hooper et al., 2013). Although this may provide insight into general trends, large-scale assessments have limited value for inferring what happens in different classrooms. Many small-scale classroom studies have provided useful and detailed information about classroom practices – but it has been hard to relate these cases to findings in large-scale assessments. Further, these small case studies are often difficult to compare with one another, making knowledge aggregation across studies of instructional practices very challenging (Klette & Blikstad-Balas, 2017). As emphasized by Grossman and McDonald (2008), educational research has traditionally favored case studies, which have poor guidelines for empirical generalization, combined with a preference for qualitative data, which is not easily combined with quantitative research on teacher effectiveness.

In our ongoing research project, Linking Instruction and Student Achievement (LISA), we aim to combine detailed data about the teaching practices happening in the classroom with both student perceptions and student achievement. Thus, the research design we will elaborate in this article is based on the premise that the educational field would benefit from juxtaposing instructional data from the macro- and meso-level (for example, achievement data or large-scale student surveys) with data from the micro-level (for example, analyses of classroom practices) (Blikstad-Balas, 2016; Klette & Blikstad-Balas, 2017; Lemke, 2007; Snell, 2011). We believe there is a need to go beyond correlational patterns found in large-scale student assessments, thereby attempting to open the "black-box" of instruction (Rowan, Camburn & Corenti 2004). Several scholars have specifically asked for studies that correlate achievement data with records of learning processes and teaching strategies (Baumert et al., 2010; Bryk et al., 2010; Kuger & Klieme, 2016). This line of research has expanded over recent years and seems quite promising (Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2009; Cohen & Grossman, 2016; Grossman, Loeb, Cohen, & Wyckoff, 2013; Raudenbush, 2008) It has also

become clear that there is a need to understand determinants at different levels (class, school, regions), and recently, advances in hierarchical linear and structural equation modeling have enabled researchers to generate theoretical constructions that adequately represent a larger complexity of interacting variables. Such a research program is quite demanding because it requires a concerted effort by researchers from different fields of expertise and traditions: psychometric research, advanced statistics, microgenetic classroom studies, and research into domain-specific instruction and learning. A main ambition of our ongoing research project is precisely to advance such integration. The project title reveals this goal: Linking Instruction and Student Achievement (LISA).

In this article, we begin by providing theoretical background and the status of knowledge relating to instructional quality. We will go on to argue that video data has shown particular promise in studies aiming to obtain systematic data from a range of classrooms in order to compare classroom practices. We then present the three key data sources and design components of the LISA project and discuss the value of combining these three data sources. Finally, we will outline some of our preliminary findings to date and point to future research possibilities.

Instructional quality – a short overview

For decades, researchers from a range of different traditions have investigated what characterizes "effective teaching" with different definitions and analytical foci – and several attempts have been made to define and conceptualize *instructional quality* (Baumert et al., 2010; Fischer & Neumann, 2012; Lipowsky et al., 2009; Raudenbusch, 2008; Vieluf & Klieme, 2011). A recent definition refers to instructional quality as "a particular pattern of instructional variables and conditions influencing instruction that positively influences students' achievement" (Fisher & Neumann, 2012, p. 116). Consensus is emerging that there are three to four basic dimensions that are critical for high-quality instruction across subjects. The dimensions *instructional clarity*, *cognitive activation*, *discourse features* and *supportive climate* were chosen to form LISA's main analytical perspective precisely because they have proven to be essential parts of instructional quality (Klette, 2015; Gersten et al. 2009; Praetorius, Pauli, Reusser, Rakoczy, & Klieme, 2014). We will elaborate these key dimensions, with a special emphasis on LISA's two target subjects, mathematics and language arts (LA), by looking at gains in test scores in numeracy and reading.

Instructional clarity: This dimension seeks to capture, for instance, the degree of explicitness in instruction, the clarity of learning goals, and how content-oriented instruction occurs. Across content areas, explicit strategies are

considered the most effective, while more implicit approaches and teaching methods such as inquiry-based methods fail to show any decisive influence on students' learning (Seidel & Shavelson, 2007). With regard to both reading and writing, a number of scholars suggest that explicit instruction of a specific strategy, including how and when to use it, is an essential component of instruction (Afflerbach, Pearson, & Paris, 2008; Duke & Pearson, 2002; Graham, Harris, & Chambers, 2016). As Graham et al. (2016) emphasize, highly effective teachers set clear goals for their students and establish high expectations. Another concept that falls under the dimension of instructional clarity, essential in both LA and mathematics instruction, is modeling, that is, showing explicitly how relevant strategies and ways of working can be applied in action (Afflerbach et al., 2008; Gersten et al., 2009; Graham et al., 2016; Ogle & Lang, 2011; Snow, Griffin, & Burns, 2005).

Cognitive activation: The quality of different tasks, cognitive challenges, and content coverage are all part of the dimension of cognitive activation. As explained by Lipowsky et al. (2009, p. 529), cognitive activation as an instructional practice "encourages students to engage in higher-level thinking and thus to develop an elaborated knowledge base." In mathematics, studies have shown the importance of developing conceptual understanding by explicitly attending to concepts and the connections between mathematical facts and ideas and mathematical representations (Hiebert & Grouws, 2007). A central principle in teaching language arts is that students must engage in a variety of reading and writing tasks with a variety of different texts (Gambrell, Malloy, & Mazzoni, 2011; Graham et al., 2016; Ogle & Lang, 2011). A key part of assessing cognitive activation across subjects is to question who is actually doing the intellectual work in the lesson and to assess the intellectual challenge students are given in their instruction. Cognitive activation increases when students are challenged, faced with contradictions, asked to reflect explicitly on their learning, and understand links between new content and previous content (Lipowsky et al., 2009). Similarly, it decreases if the instruction fits within what has been labeled "the transmission" approach to learning (Wade & Moje, 2000), whereby students are expected simply to transmit subject-matter knowledge and apply known procedures.

Discourse features: Several educational researchers argue that providing opportunities for participation in subject-related discourse is among the most important aspects of teaching (Klette, 2015; Airey, 2013; Shanahan, 2008). Although it has been difficult to establish a strong connection between interactive patterns and students' learning (Grossman, Loeb, Cohen, & Wyckoff, 2013), verbal communication and the quality of instructional conversation – including teachers' uptake and elaboration of students' ideas – play an important role in high-quality instruction (Cazden, 2001; Mortimer & Scott, 2003; Nystrand, 1997). The quality of students' participation in learning activities also relates to engagement (Skinner, Kindermann, & Furrer, 2009). Highly relevant

in language arts, Gambrell and colleagues (2011, p. 27) emphasize how several studies "provide evidence that discussions of text promote reading comprehension, motivation to read, and higher order thinking skills." Regarding the teaching of reading comprehension strategies, Block and Duffy (2008) claim there is a need to reduce the amount of teacher-dominated talk in the classroom in order to increase the amount of time students spend on authentic tasks. Further, they claim that "when instruction is too teacher dominated, students do not learn to apply the strategies" (Block & Duffy p. 28). In mathematics, Brophy (2000) makes a distinction between discourse that simply elicits short mathematical answers and more preferable discourse featuring "sustained and thoughtful development of key ideas".

Supportive climate: A supportive climate is believed to significantly influence learning outcomes (Hattie, 2009). This dimension concerns creating an environment of respect and rapport while managing classroom procedures. Classroom procedures are important because effective teachers may have more efficient routines for transitioning between activities and better classroom management, leading to more time dedicated to instruction (Lipowsky et al., 2009; Seidel & Shavelson, 2007; Wayne & Youngs, 2003). Aspects of teacher–learner interactions, such as positive and constructive teacher feedback, supportive teacher–student relationships, and general caring behavior by the teachers, are all foundations of a supportive climate (Lipowsky et al., 2009). This is important across subjects. Gambrell and colleagues (2011) rank "building a whole class context that emphasizes community and collaboration" as one of ten evidence-based practices that contribute to comprehensive literacy instruction, while a review by Graham et al. (2016) emphasizes that creating a supportive writing environment is crucial for teaching writing. Studies have also found that an engaging classroom environment can motivate students to read strategically, thus impacting their reading comprehension (Afflerbach, Cho, Kim, Crassas, & Doyle, 2013). Franke, Kazemi, and Beatty (2007) underscore the role of supportive climate for probing students' understanding and asking for explanations in mathematics.

A new generation of video studies

Scholars agree that video analysis has multiple and significant advantages in developing our understanding of teaching and learning processes (Hiebert, 2003; Janík & Seidel, 2009; Fischer & Neumann, 2012). David Clarke and colleagues argue that video recordings "…provide a much richer portrayal of classroom practices than would be possible from any single analysis" (2006a, p. 6). Drawing on video documentations from science classrooms, Fischer and Neumann (2012) claim that video studies are especially interesting for decomposing qualities in teaching, because such studies are able to capture students' and teachers' behaviors in one "package."

Video recordings reveal classroom practices more clearly, facilitate discovery of new alternatives, and stimulate discussions about pedagogical

choices within each classroom, thereby deepening educators' understanding of teaching. Video also facilitates the study of complex processes and integration of qualitative and quantitative analyses. It enables coding from multiple perspectives and new ways of communicating findings and results. Furthermore, video data can be stored in a form that allows for subsequent analyses, novel analyses, and combining data and carrying out joint analyses. Video studies have proven to be a valuable tool for investigating instruction both at the level of individual teachers and within larger study designs involving samples of teachers from a country or region as well as between countries or regions. Video analysis allows identification of subject-specific patterns of instruction and "cultural scripts" (Stigler & Hiebert, 1999). It also enables identifying cause–effect relationships in different teachinglearning scenarios (Fischer & Neumann, 2012) and allows for in-depth analyses of instructional processes (Clarke, Keitel, & Shimizu, 2006b; Klette, 2009; Borko, Jacobs, Eiteljorg, & Pittman, 2008).

The growing interest in video design can be traced to rapid development of technology that allows easy storage and online streaming. Video recording equipment is now miniaturized and portable and can be remotely controlled and operated by individual researchers or teachers themselves, thus making such studies feasible and less intrusive on the everyday life of classrooms. New technologies in this field have been paralleled by major developments in coding and processing instruments, software for analyzing video data (e.g., StudioCode, Interact, Observer XT), and systems and infrastructures that facilitate sharing of raw data and analyses (see, for example, the TeachingLearningVideoLab_oslo, http://www.uv.uio.no/ils/om/tlv_lab/) and targeted and integrative analyses.

Several challenges remain in the development of valid common protocols and coding manuals as well as search and storage facilities that will preserve contextual richness while also providing the necessary standardization for collaborative video analysis and reuse of data. Recent reviews on coding protocols (Bill and Melinda Gates Foundation [BMGF], 2012;; Bell et al. (forthcoming), Klette & Blikstad-Balas, 2017) have discussed how coding protocols might differ in analytic focus and granularity, whether they should be treated as generic versus subject-specific manuals, and how they related to views of learning. Coding protocols might also differ in "observation systems" (Bell et al., forthcoming) in terms of required training and certification, time segments for coding, and accompanying rubrics and scales for scoring. In the LISA study, we have chosen to use the Protocol for Language Arts Teaching Observation protocol (PLATO), developed by Pam Grossman (see Grossman et al., 2013) as the overarching observation system. PLATO addresses four domains that are assumed to be critical for middle and secondary school instruction: Instructional Scaffolding, Cognitive Demand, Representation and Use of Content, and Classroom Climate. Each of these domains are divided into three or four sub-elements. Although developed for LA instruction, the PLATO protocol might

also be profitable for mathematics instruction (Cohen, 2014); the present study uses PLATO for coding both mathematics and LA classrooms.

No single instrument could cover all factors critical for students' learning. We have chosen PLATO because: (a) it resonates well with existing research relevant for this kind of study; (b) the four main domains replicate areas outlined as critical for student learning in the research literature; (c) it builds on an observation system that is feasible and applicable; (d) it makes it possible to systematically compare instruction across subjects, countries and educational settings (in our own project, but also by comparing our findings to other studies drawing on PLATO, and finally (e) it provides an opportunity to test out possible cultural biases and specificities embedded in the PLATO instrument (Klette & Blikstad-Balas, 2017)

Methodological perspectives: LISA's research design

The overall sample in LISA consists of 47 mathematics classrooms and 46 LA classrooms from 49 different schools across Norway. To increase generalizability, the classrooms were sampled from different geographical areas across Norway, including urban and rural schools and schools reflecting variation in socioeconomic background. The schools were also sampled on the basis of achievement scores on the national reading tests in 2012 and 2013 as a basis for calculating gains from grade eight to grade nine to ensure that the sample included schools with high, average, and low gains. At each school, we video-recorded one eighth grade class for four consecutive lessons in mathematics and language arts, respectively. These students answered a survey concerning instructional repertoires in both subjects. We also gathered these students' achievement scores on the national tests mentioned above, which constitute the basis for calculating achievement gains from 8th to 9th grade. Informed consent was given by principals at each school, teachers in the class, students, and students' parents. The three different kinds of data from each classroom (and the potential to combine them) will be elaborated in the following.

Student survey
Solicitation of student perceptions of their teachers has a long history. However, the kind of questions that students are asked – and how these relate to research on instructional quality – is essential. In many countries, student surveys have been used somewhat narrowly, to provide information about students' general well-being, their general attitudes towards school or school subjects, and how they perceive their school climate, rather than to investigate the quality of instruction of specific teachers. While students' perception and assessment of

their learning environment has been found to be reliable and even predictive of learning outcomes (e.g., Kane & Staiger, 2012; Wallace, Kelcey, & Ruzek, 2016), the methodology used to obtain students' perceptions is crucial. In the words of Wallace et al. (2016, p. 1835), "adolescents' daily experiences of exposure to particular teacher actions that affect student outcomes must be translated into high-quality survey items."

The LISA study uses the *Tripod Student Perception Survey*, developed by Ronald Ferguson (2010). This is a thoroughly validated instrument and was the survey used in the Bill & Melinda Gates-funded study Measures of Effective Teaching (MET). The survey focuses on the instructional quality of the classroom, assessing three areas (thus, tripod): content, pedagogy, and relationships. These areas are linked to seven domains: care, control, clarify, challenge, captivate, confer, and consolidate, henceforth called the seven C's. These constructs all resonate well with research on instructional quality, as they address the four general dimensions linked to instructional quality reviewed above. The Tripod survey was developed for an American context and designed for three levels of students: grades K–2, 3–5, and 6–12. In the present study, we used a tested and implemented version of the Tripod Secondary Instrument translated into Norwegian for Norwegian students. The survey was piloted to ensure the translations were sound and the questions easy to understand. In addition to the 35 items from the Tripod (secondary version), we added three items suggested by the students participating in the pilot as essential information about instructional quality: "The teacher provides instruction too fast," "The teacher understands when students are tired or have had a really long day," and "The teacher spends time helping all students." The students responded to each item on a five-point scale: 1 = never, 2 = seldom, 3 = sometimes, 4 = often, and 5 = always.
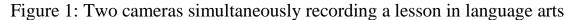
Classroom observations

As previously described, video recordings are particularly valuable in classroom analysis specifically due to the possibility of systematically investigating complex educational settings and providing opportunities for deconstructing qualities in teaching (Klette, 2009; Blikstad-Balas, 2016; Heath, Hindmarsh, & Luff, 2010; Snell, 2011). The video design used to record the lessons in the LISA study relies on two cameras simultaneously recording the same lesson: one capturing the class and one focusing on the teacher (see Figure 1).

Camera capturing the class   Camera capturing the teacher

Figure 1: Two cameras simultaneously recording a lesson in language arts

We used two microphones, one placed on the teacher and one fixed microphone in the middle of the room capturing the class. This made it possible to obtain reasonably good audio for both whole-class discourse and teacher interaction with one or a few students at a time. We recorded four consecutive lessons in mathematics and language arts across the 49 classrooms because observation studies from K–12 classrooms suggest that four consecutive lessons per classroom provides sufficient information to obtain a first overview of teaching quality (Ball & Hill, 2009; Bill & Melinda Gates Foundation, 2012; Klette, 2009).

The video-recorded lessons were coded using PLATO. PLATO focuses on 13 elements of instruction that are highlighted in existing literature on effective instruction in secondary LA education (Cohen & Grossman, 2016). We would like to emphasize that even though PLATO was designed to measure instruction in language arts, it has been successfully used to analyze mathematics (Cohen, 2014, 2015; Cohen & Brown, 2016), and in the present study we have therefore used it to analyze both LA instruction and mathematics instruction. Aside from one PLATO-element, text-based instruction, which aims to evaluate the use and production of authentic texts in a lesson, we have found the PLATO categories to be equally relevant in mathematics and LA. The dimensions in PLATO resonate well with the four domains of instructional quality previously described in our review. PLATO was also a part of the previously mentioned MET study. PLATO is a relevant tool to analyze the video observations in the present study; (a) to examine and structure observations of teaching in a systematic way, (b) to use a thoroughly validated protocol that provides a consistent lens for looking at different aspects associated with effective instruction across lessons in each of the 49 classrooms, and (c) to make interpretations explicit, traceable, and measurable (Klette, 2009). Further, using standardized observation manuals such as PLATO enables comparison and knowledge aggregation beyond what is possible in single case studies (Klette & Blikstad-Balas, 2017). Each recorded lesson was divided into events of 15 minutes and the events were coded by at least one certified PLATO rater. The rater coded all the 13 dimensions of

PLATO using a research-based theory-driven four-point scale that distinguishes between no evidence and solid evidence of a given teaching practice. PLATO differentiates between high-end and low-end evidence of instructional practices – and makes nuances between these categories explicit. Rather than coding whether or not a teacher uses, for example, modelling in his or her teaching, PLATO differentiates between degrees of modelling. As Klette & Blikstad-Balas (2017) argue, this differentiation illustrates a shift from discussing modelling per se to discussing the differences between no modelling, partially successful modelling and higher-level evidence of successful modelling.

Achievement data from national tests
We gathered achievement data from the national tests (Norw. "nasjonale prøver") in reading and numeracy for each class participating in the study. All students in eighth and ninth grade take the same national tests in reading and numeracy each year in order to track students' progress/gains over a school year. Thus, we can compare the average gains for each classroom with the national average. Further, the average gains in numeracy and reading can be linked to PLATO scores and student survey scores in mathematics and language arts within and across classrooms. The three datasets form the basis for statistical analyses studying the relationship between instruction, perceptions, and student achievement.

What have we found so far?

The LISA project is ongoing, and we have not yet reported our main findings regarding links between instruction and student achievement. However, we share here some interesting preliminary findings from classroom observations and student feedback measures.

Initial findings from the survey data show significant differences between classrooms in terms of students' overall rating of their teachers' instruction. In general, students across all classrooms in both mathematics (n=991) and LA (n=964) consistently rated their teachers highest on the three items reported in Table 1. The three items in Table 2 are consistently scored the lowest for both subjects.

Table 1 Mean scores in Mathematics and LA for the three highest rated survey items

	Mathematics (N=991)		Language Arts (N=964)	
	Mean	SD	Mean	SD
If you don't understand something, my teacher explains it another way.	4.30	0.90	4.19	0.91
Students in this class treat the teacher with respect.	4.30	0.75	4.23	0.79
The teacher spends time helping all students.	4.25	0.88	4.10	0.94

Table 2 Mean scores in Mathematics and LA for the three lowest rated survey items

	Mathematics (N=991)		Language Arts (N=964)	
	Mean	SD	Mean	SD
Students get to decide how activities are done in this class.	2.71	0.99	2.76	1.00
Students speak up and share their ideas about class work.	3.14	1.06	3.19	1.11
My teacher makes learning enjoyable.	3.36	1.09	3.13	1.09

Interestingly, students rate their math teachers higher than their LA teachers for most items. However, we found no significant gender differences of student respondents for any item, which indicates that possible gender differences in terms of attitudes towards the subjects have not affected students' scorings. For the six items recognized as key for student achievement by the MET study (Kane & Staiger, 2012), the mean scores in mathematics, as well as language arts, were higher than the total average for all 38 items, see Table 3.

Table 3 Mean scores in Mathematics and LA for six survey items recognized as key for student achievement by the MET study and total mean score for all items

	Mathematics (N=991)		Language Arts (N=964)	
	Mean	SD	Mean	SD
Students in this class treat the teacher with respect.	4.30	0.75	4.23	0.79
My classmates behave the way my teacher wants them to.	3.85	0.78	3.78	0.80
Our class stays busy and doesn't waste time.	3.91	0.80	3.89	0.75
In this class, we learn a lot every day.	3.95	0.84	3.81	0.92
In this class, we learn to correct our mistakes.	3.97	0.90	3.79	0.95
My teacher explains difficult things clearly.	4.09	0.91	3.99	0.87
Mean score for all items.	3.80	0.53	3.75	0.55

This is promising, because the MET study found these items to be particularly important for improving student achievement.

Tentative analyses from the video data indicate differences and similarities between mathematics and LA classrooms with regard to the distribution of the four score codes (1–4) across the various elements of instructional quality measured by PLATO. The scores are quite similarly distributed in mathematics and LA lessons for the PLATO elements Purpose, Connection to Prior Knowledge, Feedback, Classroom Discourse, and the two elements capturing classroom climate, Behavioral Management and Time Management. However, we find a higher percentage of high scores (score 3 and score 4) in mathematics than in language arts for the elements Intellectual Challenge, Representation of Content, Modeling, and Strategy Use and Instruction.

We also found that Behavioral Management and Time Management receive consistently high scores (score 3 and score 4) in the observed classrooms in both mathematics and language arts. Over 90% of the video segments are scored at the high end for Behavioral Management, while equal scores in the high end for Time Management are over 80%. Our study thus shows a very different picture than the commonly portrayed idea of lower secondary schools having severe disciplinary problems. This finding is also consistent with our survey results, where the overall scores on constructs "care" and "control" suggest respectful and caring teachers who manage their classes well.

Our preliminary analyses further suggest that few lesson segments score at the high end of important PLATO elements for scaffolding techniques and making learning aims visible, such as Purpose, Connection to Prior Knowledge, Feedback, and Explicit Strategy Use and Instruction. The element Strategy Use and Instruction shows interesting differences between mathematics and LA classrooms. Mathematics teachers in our material seem to use explicit strategy instruction somewhat more frequently than their LA colleagues. This is also the case for the element Modeling. These differences across subjects will be studied in more detail in future publications, where the differences also will be examined in relation to both student surveys and achievement gains.

Several Master's students assisting with the LISA study have also conducted smaller studies on the same video material, directly relevant to the LISA ambitions. Some of them will be mentioned here to illustrate how qualitative small-scale studies can add to and expand our understanding on trends in the overall material. Marie Dåsvatn investigated what characterizes the LA teachers in the overall sample scoring highest in the PLATO category Feedback. Her qualitative analyses illustrate how teachers who all score in the high end on the category measuring feedback have different instructional practices. Thus, while these teachers have in common that they are all able to provide their students with high-quality feedback, their ways of doing so differ (Dåsvatn, 2016). Another MA thesis on feedback by Sofie Jensen (2017) assessed to what degree LA teachers provide feedback during whole class teaching, and to what degree the feedback provided align with the ideals of assessment for learning. She found that while there were clear learning goals in the LA lessons, and teachers

did provide a lot of feedback, this feedback was very often short and vague. Very few teachers provided substantial feedforward. Alexander J. V. Selling (2017) investigated teachers' use of learning goals in 18 mathematics lessons. He found that a majority of the learning goals were procedural and associated with specific tasks and activities. He also found that these goals were often framed in a very short time span, and that few teachers talked about how mathematical knowledge could be used in other contexts than the specific tasks. Finally, Maren Liestøl Jahnsen (2017) investigated the classroom discourse in whole class discussions in five different classrooms, and analyzed whether the classroom discourse could be labeled as dialogic teaching. She found that one of the classrooms had a dialogic teaching approach, while the others were largely characterized by the traditional Initiation Response Evaluation (IRE/F) patterns. These smaller studies all show the importance of in-depth studies of instructional quality to obtain fine-grained knowledge.

Looking forward: Implications

Previous research on instruction has revealed complex and nontrivial relationships between instructional variables – including school and teacher characteristics, teacher cognition and beliefs, teachers' and students' activities during instruction, and last but not least, learning outcomes. As we have argued, it is crucial to look at instructional quality from different angles and to design studies that link data from different levels (micro-meso and-macro) to investigate how teachers' instruction relates to students' achievement, as well as students' perception of instructional quality.

    In our ongoing study, systematic video analyses combined with student perception surveys provide reliable and qualified information about the merits of different teaching practices. Our findings indicate that students differentiate between a positive view on the concrete enactment of teaching in subjects and their personal general attitudes toward each subject (Authors, in progress). Video observations made with a validated observation manual are providing new and comparable insights into patterns of instruction in both mathematics and language arts, as well as interesting differences between these subjects. Video analyses seem to allow differentiation between high-quality and low-quality teaching, but they also require more in-depth analyses together with integrative multi-level analyses.

About the Authors

Kirsti Klette is Professor at the Department of Education and School Research, University of Oslo. Her research interests are classroom research and instructional quality. She is currently leading the large scale classroom video study Linking Instruction and Student Achievement (LISA),and the eInfrastructure for video research (eVIR) project.
Institutional affiliation: Department of teacher education and school research, University of Oslo, P.O. box 1099 Blindern, 0317 Oslo
E-mail: Kirsti.klette@ils.uio.no

Marte Blikstad-Balas is Associate Professor at the Department of Education and School Research, University of Oslo. Her research interests are literacy and how texts are used in school across subjects. She is currently a part of the large scale classroom video study Linking Instruction and Student Achievement (LISA).
Institutional affiliation: Department of teacher education and school research, University of Oslo, P.O. box 1099 Blindern, 0317 Oslo
E-mail: marte.blikstad-balas@ils.uio.no

Astrid Roe is Research Professor at the Department of Education and School Research, University of Oslo. Her research interests are reading and writing instruction across subjects, assessment of reading and the use of test results to improve students' reading literacy. She is part of the large scale classroom video study Linking Instruction and Student Achievement (LISA).
Institutional affiliation: Department of teacher education and school research, University of Oslo, P.O. box 1099 Blindern, 0317 Oslo
E-mail: astrid.roe@ils.uio.no

References

Afflerbach, P., Cho, B., Kim, J. Y., Crassas, M. E., & Doyle, B. (2013). Reading: What else matters besides strategies and skills? *The Reading Teacher*, 66(6), 440–448.

Afflerbach, P., Pearson, D., & Paris, S. (2008). Clarifying differences between reading skills and reading strategies. *The Reading Teacher, 61*(5), 364–373.

Airey, J. (2013). Disciplinary literacy. In E. Lundqvist, R. Säljö, & L. Östman (Eds.), *Scientific literacy. Teori och praktik* (pp. 41–58). Malmö: Gleerups.

Ball, D. L., & Hill, H. C. (2009). Measuring teacher quality in practice. In D. H. Gitomer (Ed.), *Measurement issues and assessment for teaching quality* (pp. 80–98). Thousand Oaks, CA: Sage Publications.

Bell, C. , Dobbelaer, M., Klette, K. Visscher A. (under review). Qualities of Classroom Observation systems. *School Effectiveness and School Improvement Journal*

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., . . . Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal, 47*(1), 133–180.

Bill & Melinda Gates Foundation (BMGF) (2012). Gathering feedback for teaching: Combining high

quality observations with student surveys and achievement gains. Seattle, WA: The Bill & Melinda Gates Foundation.

Blikstad-Balas, M, (2016) Key challenges of using video when investigating social practices in education:Contextualization, magnification, and representation. International Journal of Research & Method in Education. Epub ahead-of-print: 1–13. DOI: 10.1080/1743727X.2016.1181162.

Block, C. C., & Duffy, G. G. (2008). Research on Teaching Comprehension: Where we've been and where we're going. In C. C. Block & S. R. Parris (Eds.), *Comprehension instruction. Research-based best practices*. New York: The Guilford Press.

Borko, H., Jacobs, J., Eiteljorg, E. & Pittman, M.E. (2008). Video as a tool for fostering productive discussions in mathematics professional development. *Teaching and Teacher Education, 24*, 417–436.

Boyd, D. J., Grossman, P. L., Lankford, H., Loeb, S., & Wyckoff, J. (2009). Teacher preparation and student achievement. *Educational Evaluation and Policy Analysis, 31*(4), 416–440.

Brophy, J. E. (2000). *Teaching*. UNESCO Educational Practices Series -1. International Academy of Education and the International Bureau of Education (IAE) http://www.ibe.unesco.org.

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

Cazden, C. B. (2001). *Classroom discourse: The language of teaching and learning*. London, UK: Heinemann.

Cohen J. (2014).Challenges in identifying high leverage practices. PhD Thesis. *University of Virginia,   Curry School of Education.*

Cohen, J. (2015). The challenge of identifying high-leverage practices. *Teachers College Record*,

117(8), 1–41.

Cohen, J., & Brown, M. (2016). Teaching quality across school settings. *The New Educator*, 12(2),

191–218.

Cohen, J., & Grossman, P. (2016). Respecting complexity in measures of teaching: Keeping students and schools in focus. *Teaching and Teacher Education, 55*, 308–317.

Clarke, D., Emanuelsson, J., Jablonka, E. & Chee Mok, I. A. (Eds.), (2006a). *Making connections: Comparing mathematics classrooms around the world.* Rotterdam: Sense Publishers.

Clarke, D., Keitel, C., & Shimizu, Y. (Eds.) (2006b). *Mathematics Classrooms in Twelve Countries: The Insider's Perspective*. Rotterdam: Sense Publishers.

Duke, N. K., & Pearson, P. D. (2002). *Effective practices for developing reading comprehension*. In

A..Farstrups & S.J. Samuels (Eds.) What Research Has to Say About Reading Instruction, Third Edition - © 2002, International Reading Association (pp 205- 242)

Dåsvatn, M. (2016). *Muntlige tilbakemeldinger i norskfaget - En videostudie av fem klasserom [Oral feedback in Norwegian language arts - a video study of five classrooms]. Master thesis.* University of Oslo, Oslo.

Ferguson, R. (2010). *Student perceptions of teaching effectiveness. Discussion brief.* Cambridge, MA: National Center for Teacher Effectiveness and the Achievement Gap Initiative, Harvard University.

Fischer, H. E., & Neumann, K. (2012). Video analysis as a tool for understanding science instruction. In D. Jorde & J. Dillon (Eds.), *Science education research and practice in Europe. Retrospective and prospective* (pp. 115–139). Rotterdamn, NL: Sense Publisher.

Franke, M.L., Kazemi, E. and Beatty, D. (2007). Mathematics teaching and classroom practice. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 225–256). Greenwich, CT: Information Age Publishing.

Gambrell, L. B., Malloy, J. A., & Mazzoni, S. A. (2011). Evidence-based best practices in comprehensive literacy instruction. In L. M. Morrow & L. B. Gambrell (Eds.), *Best practices in literacy instruction*. New York: The Guilford Press.

Gersten, R., Chard, D. J., Jayanthi, M., Baker, S. K., Morphy, P., & Flojo, J. (2009). Mathematics instruction for students with learning disabilities: A meta-analysis of instructional components. *Review of Educational Research, 79*(3), 1202–1242.

Graham, S., Harris, K. R., & Chambers, A. B. (2016). Evidence-based practice and writing instruction: A review of reviews. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (Vol. 2, pp. 211–226). New York & London: The Guilford Press.

Grossman, P., Loeb, S., Cohen, J., & Wyckoff, J. (2013). Measure for measure: The relationship between measures of instructional practice in middle school English language arts and teachers' value-added scores. *American Journal of Education, 119*(3), 445–470.

Grossman, P., & McDonald, M. (2008). Back to the future: Directions for research in teaching and teacher education. *American Educational Research Journal, 45*(1), 184–205.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.

Heath, C., Hindmarsh, J., & Luff, P. (2010). *Video in qualitative research: Analysing social interaction in everyday life*. Los Angeles, CA: Sage.

Hiebert, J. (2003). *Teaching mathematics in seven countries: Results from the TIMSS 1999 video study.* Collingdale, PA: DIANE Publishing.

Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. *Second handbook of research on mathematics teaching and learning, 1*, 371–404.

Hooper, M., Mullis, I. & Martin, M. (2013). TIMSS 2015 Context Questionnaire Framework., In Mullis, I.V.S. & Martin, M.O. (Eds.). (2013). TIMSS 2015 Assessment Frameworks. pp. 61-82.

Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Jahnsen, M. L (2017). Det dialogiske klasserommet – finnes det? En videostudie av fem læreres tilnærming til dialogisk undervisning. [Dialogic teaching – does it exist? A video study of five teachers' approach to dialogic teaching]. Master thesis. University of Oslo.

Janík T., & Seidel T (eds) The Power of Video Studies in Investigating Teaching and Learning in the Classroom. Münster: Waxmann Publishing.

Jensen, S. (2017). Norsklæreres bruk av underveisvurdering i muntlig kontekst. En videostudie av fem klasserom [Norwegian LA teachers use of assessment for learning in an oral context. A video study of five classrooms]. Master thesis. University of Oslo.

Kane, T.J.& Staiger, D.O (2012) Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains. Research paper, MET Project, Bill & Melinda Gates Foundation.

Klette, K. (2009) Challenges in strategies for complexity reduction in video studies. Experiences from the PISA+ study. In: Janik T and Seidel T (eds) The Power of Video

Studies in Investigating Teaching and Learning in the Classroom. Münster: Waxmann Publishing, pp.61–83.

Klette, K. (2015) Introduction: Studying interaction and instructional patterns in classrooms. In: Klette K, Bergem OK and Roe A (eds) Teaching and Learning in Lower Secondary Schools in the Era of PISA and TIMSS. New York, NY: Springer, pp.1–14.

Klette, K. & Blikstad-Balas, M. (2017) Observation manuals as lenses to classroom teaching: Pitfalls and possibilities. European Educational Research Journal. Ahead-of-print: april 2017, doi 10.1177/1474904117703228

Konstantopoulos, S., & Chung, V. (2011). The persistence of teacher effects in elementary grades. *American Educational Research Journal, 48*(2), 361–386.

Kuger, S., & Klieme, E. (2016). Dimensions of context assessment. In S.Kuger, E Klieme**,** N. Jude, D. Kaplan**,** **(**Eds.) *Assessing Contexts of Learning. An International Perspective* (pp. 3–37): Springer.

Lemke, J. (2007). Video epistemology in-and-outside the box: Traversing attentional spaces. In R. Goldman, R. D. Pea, B. Barron, & S. J. Derry (Eds.), *Video research in the learning sciences* (pp. 39–51): Lawrence Erlbaum.

Lipowsky, F., Rakoczy, K., Pauli, C., Drollinger-Vetter, B., Klieme, E., & Reusser, K. (2009). Quality of geometry instruction and its short-term impact on students' understanding of the Pythagorean Theorem. *Learning and Instruction, 19*(6), 527–537.

Mortimer, E., & Scott, P. (2003). *Meaning making in secondary science classrooms.* London, UK: McGraw-Hill Education.

Nystrand, M. (1997). *Opening dialogue: Understanding the dynamics of language and learning in the English classroom.* New York, NY: Teachers College Press.

OECD (2010). *PISA 2009 Results (Volume II): Overcoming Social Background – Equity in Learning Opportunities and Outcomes* Paris: OECD Publications.

OECD (2016): *PISA 2015 results: (Volume l) excellence and equity in education.* Paris: OECD Publications.

Ogle, D., & Lang, L. (2011). Best practices in adolescent literacy instruction. In L. M. Morrow & L. B. Gambrell (Eds.), *Best practices in literacy instruction* (pp. 127–157). New York: The Guilford Press.

Praetorius, A.-K., Pauli, C., Reusser, K., Rakoczy, K., & Klieme, E. (2014). One lesson is all you need? Stability of instructional quality across lessons. *Learning and Instruction, 31*, 2–12.

Raudenbush, S. W. (2008). Advancing educational policy by advancing research on instruction. *American Educational Research Journal, 45*(1), 206–230.

Rowan, B. E., Camburn, E, & Correnti, R. (2004). Using teacher logs to measure the enacted curriculum in large scale surveys. Insights from the Study of Instructional Improvement. *Elementary School Journal*, Vol. 105, 75 -102.

Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research, 77*(4), 454–499.

Selling, A. J. V (2017). Målsättningar i matematik [Learning goals in mathematics]. Masther thesis. University of Oslo.

Shanahan, T. a. S., C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content-area literacy. *Harvard Educational Review, 78*(1), 40–59.

Skinner, E. A., Kindermann, T. A., & Furrer, C. J. (2009). A motivational perspective on engagement and disaffection: Conceptualization and assessment of children's behavioral and emotional participation in academic activities in the classroom. *Educational and Psychological Measurement, 69*(3), 493–525.

Snell, J. (2011). Interrogating video data: Systematic quantitative analysis versus micro-ethnographic analysis. *International Journal of Social Research Methodology, 14*(3), 253–258.

Snow, C. E., Griffin, P., & Burns, M. S. (2005). *Knowledge to support the teaching of reading: Preparing teachers for a changing world*. San Francisco, CA: Jossey-Bass.

Stiegler, J. W. & Hiebert, J. (1999). *The teaching gap: Best ideas from the world's teachers for*
improving education in the classroom. New York: Free Press.

Timperley, H., & Alton-Lee, A. (2008). Reframing teacher professional learning: An alternative policy approach to strengthening valued outcomes for diverse learners. *Review of Research in Education, 32*(1), 328–369.

UNESCO. (2017). *Education for sustainable development goals. Learning objectives*. Paris, France: UNESCO. Retrieved from http://unesdoc.unesco.org/images/0024/002474/247444e.pdf

Vieluf, S., & Klieme, E. (2011). Cross-nationally comparative results on teachers' qualification, beliefs, and practices *Expertise in mathematics instruction* (pp. 295–325). New York, NY: Springer.

Wade, S. E., & Moje, E. B. (2000). The role of text in classroom learning. In M. L. Kamil, P. B. Mosenthal, D. B. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 609–627). Mahwah NJ: Erlbaum.

Wallace, T. L., Kelcey, B., & Ruzek, E. (2016). What can student perception surveys tell us about teaching? Empirically testing the underlying structure of the tripod student perception survey. *American Educational Research Journal, 53*(6), 1834–1868.

Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research, 73*(1), 89–122.