TECHNISCHE UNIVERSITÄT MÜNCHEN
TUM School of Education
Friedl Schöller-Stiftungslehrstuhl für Pädagogische Psychologie

---

The Interplay of Teacher Judgment Accuracies and Student Characteristics in Classroom Discourse

---

Maralena Veronika Weil

Vollständiger Abdruck der von der Fakultät *TUM School of Education* der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Philosophie (Dr. phil.)

genehmigten Dissertation.

Vorsitzende:	Prof. Dr. Claudia Nerdel
Prüfende der Dissertation:	1. Prof. Dr. Christina Seidel
	2. Prof. Dr. Doris Holzberger

Die Dissertation wurde am 12.06.2019 bei der Technischen Universität München eingereicht und durch die Fakultät *TUM School of Education* am 15.07.2019 angenommen.

Acknowledgements

First, I want to thank the teachers and students who participated in the studies. Thanks to my research team and our student assistants for their engagement in data collection and analysis. I also want to thank the German Research Foundation who funded the two projects in which the studies of this dissertation were settled (DFG, SE 1397/7-1 and DFG, SE 1397/5-3).

To my supervisor Prof. Dr. Tina Seidel, I want to express my deep gratitude. Without her guidance, support, encouragement and fostering my ideas this dissertation would not have been possible and the learning process would not have been so rich and enjoyable.

In addition, I would like to thank my second supervisor Prof. Dr. Alexander Gröschner for his support and feedback throughout my learning process and for all the learning opportunities.

My special thanks also to my mentors Dr. Sina Huber and Dr. Ann-Kathrin Schindler for supporting many steps of this dissertation, for providing me always constructive feedback in my learning process and always having an open ear for me with valuable advises.

Moreover, I want to thank Prof. Dr. Doris Holzberger for being the second examiner and Prof. Dr. Claudia Nerdel for being the chairperson of the examination committee. I am very grateful to them for dedicating their valuable time, supporting these final steps of my dissertation.

Additionally, I am grateful to Prof. Dr. Hilda Borko, Prof. Dr. Jonathan Osborne and Ass. Prof. Dr. Florencia Gomez Zaccarelli for the valuable insights into their projects and dialogues about our projects and my dissertation and the unforgettable time during my research stay at the Stanford Graduate School of Education.

Further, I would like to thank all my colleagues and friends who supported me on this journey and truly made this time memorable.

My deepest appreciation belongs to my husband and my family for always believing in me, supporting and encouraging me throughout this and many more journeys. This dissertation would not have been possible without them.

Table of Contents

Abstract

Throughout educational systems, classroom discourse is the main teaching method. An appreciation of student contributions and providing feedback to students are key features of productive classroom discourse. In order to improve teachers' classroom discourse practices, teacher professional development is needed (e.g. O'Connor, Michaels, Chapin, & Harbaugh, 2017). For classroom discourse student characteristics also play a central role. Especially, students' self-concept of ability is a crucial predictor for verbal contributions students make in teacher-student interactions (e.g. Jurik, Gröschner, & Seidel, 2014). To adjust instruction and support individual learning, teachers have to accurately judge student characteristics such as pre-achievement and self-concept of ability as these are relevant pre-requisites for learning (Corno, 2008). This dissertation combines different data sources, student characteristics (pre-achievement and self-concept of ability), teacher judgment accuracy of these student characteristics and video data of classrooms to investigate teacher and student interactions. Furthermore, this dissertation combined two research design approaches: In Study I, it investigated via multi-level analyses if an influence of students' characteristics and teachers' judgment accuracy exists and if so, to what extent. In Study II, it investigated in-depth, in a case study of three low-performing teachers, how a video-based TPD program can change these teachers' classroom discourse performance and which consequences from the change can be observed for student self-concept of ability and the individual teacher judgment accuracy of students' self-concept of ability. First, this dissertation showed that student pre-achievement and especially self-concept of ability were predictive of verbal teacher-student interactions. Second, adaptive teacher questioning behavior was found, but no adaptive behavior with regard to teachers' feedback. Third, all three low-performing teachers showed positive changes with regard to their classroom discourse performance by the end of the school year after participating in the Dialogic Video Cycle. Fourth, especially, students with an initially low self-concept of ability benefited from the changes in the classroom discourse as the teachers seemed to have gained a greater awareness of them, indicated by higher teacher judgment accuracies for this subgroup. The findings of this dissertation gave first insights into the possibly adaptive function of teacher judgment accuracy and emphasizes the relevance of student self-concept of ability. Furthermore, case analyses seem to be promising for further investigations in classroom research. Additionally, first hints are given that TPD programs might also have 'hidden-effects'.

Zusammenfassung

In den meisten Bildungssystemen ist das Unterrichtsgespräch ist die dominierende Lehrmethode. Wertschätzung von Schülerbeiträgen sowie konstruktives Feedback sind zentrale Merkmale eines produktiven Unterrichtsgesprächs. Um die Unterrichtsgesprächsführungskompetenz von Lehrkräfte zu verbessern, sind Fortbildungsprogramme erforderlich (O'Connor et al., 2017). Für das Unterrichtsgespräch spielen auch Schülermerkmale eine zentrale Rolle. Insbesondere das Fähigkeitsselbstkonzept ist ein entscheidender Prädiktor für die verbale Beteiligung von Schülerinnen und Schülern (z.B. Jurik et al., 2014). Um Unterricht adaptiv zu gestalten und das individuelle Lernen zu fördern, müssen Lehrkräfte die Merkmale ihrer Schülerinnen und Schüler, wie Vorwissen und Fähigkeitsselbstkonzept als Voraussetzungen für erfolgreiches Lernen akkurat einschätzen können (Corno, 2008). Diese Dissertation kombiniert verschiedene Datenquellen, Schülereigenschaften (Vorleistung und Selbstverständnis der Fähigkeiten), die Richtigkeit der Beurteilung durch den Lehrer hinsichtlich dieser Schülereigenschaften und Videodaten von Klassenzimmern, um die Interaktionen zwischen Lehrern und Schülern zu untersuchen. Des Weiteren werden in dieser Dissertation zwei Forschungsdesigns kombiniert: In Studie I wurde anhand einer Mehrebenen-Analyse untersucht, ob und in welchem Umfang ein Einfluss der Schülermerkmale und der Urteilsakkuratheit der Lehrkräfte auf das Unterrichtsgespräch vorliegt. Außerdem wurde in einer Fallstudie von drei systematisch ausgewählten Lehrkräften eingehend untersucht, wie ein videobasiertes Fortbildungsprogramm die Unterrichtsgesprächsperformanz dieser Lehrkräfte verändern kann und welche Konsequenzen sich aus den Veränderungen für das Fähigkeitsselbstkonzept der Schülerinnen und Schüler ergeben, sowie auf die individuelle Urteilsakkuratheit der Lehrkräfte in Bezug auf das Fähigkeitsselbstkonzept. In der Dissertation konnte gezeigt werden, dass erstens das Vorwissen und vor allem das Fähigkeitsselbstkonzept der Schülerinnen und Schüler verbale Lehrer-Schüler-Interaktionen vorhersagen. Zweitens, die Urteilsakkuratheit der beiden untersuchten Schülermerkmale selbst war kein Prädiktor für verbale Lehrer-Schüler-Interaktionen, es wurde ein adaptives Frageverhalten der Lehrkräfte festgestellt, jedoch kein adaptives Verhalten in Bezug auf das Feedbackverhalten der Lehrkräfte. Drittens, alle drei Lehrkräfte zeigten zum Ende des Schuljahres nach der Teilnahme am Dialogischen Video Zirkel positive Veränderungen in Bezug auf ihre Unterrichtsgesprächsperformanz. Viertens, vor allem Schülerinnen und Schüler mit einem anfänglich niedrigen

Fähigkeitsselbstkonzept profitierten von den Veränderungen im Unterrichtsgespräch, vermutlich da die Lehrkräfte ein größeres Bewusstsein für sie gewonnen hatten, welches sich durch akkuratere Einschätzungen für diese Subgruppe äußerte. Die Ergebnisse dieser Dissertation lieferten erste Einblicke in eine möglicherweise adaptive Funktion der Lehrereinschätzungen und unterstreichen die Relevanz des Fähigkeitsselbstkonzepts der Schülerinnen und Schüler. Außerdem scheinen Fallanalysen vielversprechend für weitere Untersuchungen in der Unterrichtsforschung zu sein. Des Weiteren sollte in der Forschungslandschaft mehr Bewusstsein über mögliche „Zusatzeffekte“ von Fortbildungsprogrammen entwickelt werden.

1 Introduction

1.1 Research Framework

Teaching and learning processes in classrooms are determined, to a large extent, by verbal interactions between teachers and students (Furtak & Kunter, 2012; Jurik et al., 2014; Mercer, 2008, 2010; Michaels, O'Connor, & Resnick, 2008; Pehmer, Gröschner, & Seidel, 2015a; Walshaw & Anthony, 2008). Therefore, classroom discourse is the predominant teaching method in many different countries (Mercer & Dawes, 2014; Seidel & Prenzel, 2006). Classroom discourse is settled on three main components, also known as the IRF (initiation – response – feedback) pattern. In its current way, the IRF patterns are negatively associated because it stands for a very tight ad strongly guided way of conversation. These three patterns are also the key elements for a successful and productive talk. These kinds of verbal teacher-student interactions occur very frequently in the context of school teaching and learning (Mercer, 2008, 2010). The way each pattern is practiced matters. If performed well and in the way as outlined above, they can also explain successful teaching and learning processes in classrooms and, in the long-term, positive learning developments (Chi, 2009; Mercer, Wegerif, & Dawes, 1999; Webb et al., 2014). For a successful classroom discourse first, it is important for teachers to have many students actively involved in a meaningful way (Lipowsky et al., 2009; Pauli & Lipowsky, 2007). Second, it is important for teachers to influence these verbal interactions by varying the type of questions they pose to students. It is important that not only questions that elicit the reproduction of facts are included but also questions that invite elaboration and deep thinking (Jurik et al., 2014; Oliveira, 2010; Pehmer et al., 2015a). Third, teachers need to react to student responses. In this context, the way teachers provide feedback with regard to the students' learning processes is particularly relevant (Hattie & Timperley, 2007; Jurik et al., 2014; Timperley, 2013). There are so called elements of dialogic teaching which can help to reach a more open and productive classroom discourse. One element is that classroom discourse has to be goal-oriented and structured (Resnick, Michaels, & O'Connor, 2010; Seidel & Prenzel, 2006). Teachers have to encourage students to elicit their ideas, thoughts and opinions (Chi, 2009; Oliveira, 2010). Another element is a warm and welcoming learning atmosphere where student answers are treated as a resource for the ongoing discourse (Gomez Zaccarelli,

Schindler, Borko, & Osborne, 2018). Room for mistakes and their appreciation instead of pure evaluation are part of such a learning atmosphere (Grassinger, Scheunpflug, Zeinz, & Dresel, 2018), as well as feedback containing information on students' further learning processes (Hattie & Timperley, 2007). Through implementing these elements of dialogic teaching an IRF pattern based classroom discourse, productive classroom discourse can be created (Michaels & O'Connor, 2012). However, open and productive classroom discourse is characterized by opportunities for teachers to get to know their students better and develop a greater understanding of students ideas, pre-knowledge and also judgments of their individual characteristics (Borko, Roberts, & Shavelson, 2008). But these opportunities often remain unexploited in the typical occurring closed teacher-centered IRF patterns of classroom discourse (Howe & Abedin, 2013; Pehmer et al., 2015a). It is often a challenge for teachers to change their routines and develop ways of opening classroom discourse (Seidel, 2006). Therefore, several professional development programs have been developed to help teachers improve classroom discourse (e.g., "CamTalk", van de Pol, Brindley, & Higham, 2017; "Accountable Talk", Michaels et al., 2008; "Lesson Study", Vrikki, Warwick, Vermunt, Mercer, & van Halem, 2017). Teacher professional development (TPD) provides teachers with the opportunity to reflect upon their teaching practices and routines, and to learn about their students. In the process of opening classroom discourse and making it more productive, teachers are provided with additional opportunities to get to know and understand their students. It is necessary for students to more actively participate in classroom discourse (Chi, 2009; Mercer et al., 1999; Resnick et al., 2010; Webb, 2009), and when they do, 'windows' are opened for teachers to learn about students' thinking processes and other important characteristics, such as their pre-knowledge, interests, or beliefs about their own capabilities (Glogger-Frey, Deutscher, & Renkl, 2018; Huber & Seidel, 2018; Pielmeier, Huber, & Seidel, 2018). A variety of TPD approaches have shown potential for supporting teachers in making positive changes to their interaction patterns to achieve more productive classroom discourse (Michaels et al., 2008; van de Pol et al., 2017; Vrikki et al., 2017). Especially for teachers who perform typical closed teacher-centered IRF practices intensively, the changes in their perceptions of students are generally the most strongly observable, and it is therefore worthwhile to look at such 'low-performing' teachers more closely. In order to interact adaptively with diverse students, teachers require skills to accurately judge relevant student characteristics (Corno, 2008; Machts, Kaiser, Schmidt, & Möller, 2016; Praetorius, Berner, Zeinz, Scheunpflug, & Dresel, 2013; Seidel & Shavelson, 2007), and inaccurate judgments have been shown to be

negatively related to learning and achievement gains (Anders, Kunter, Brunner, Krauss, & Baumert, 2010). It has to be taken into consideration that teachers are fairly accurate in judging student pre-achievement levels (Praetorius et al., 2013). However, it is harder for them to judge student self-concept of ability (Praetorius et al., 2013; Spinath, 2005; Südkamp & Praetorius, 2017). Teacher diagnostic skills, for example, with regard to teacher judgment accuracy, are currently gaining renewed attention in teaching and teacher research (Borko, Roberts et al., 2008; Südkamp & Praetorius, 2017). Especially, the investigation of adaptive functions in the process of teaching and with regard to addressing individual students needs is requested. Regarding the teacher judgment of individual student characteristics as well as their interplay, there are a number of studies that have been published in the last few years (Praetorius et al., 2013; Praetorius, Karst, Dickhäuser, & Lipowsky, 2011). In addition, elements of verbal interactions between teachers and students have been identified in teaching research, which might be relevant with regard to adaptive functions of teacher judgment accuracy (Mercer, 2008, 2010; Michaels et al., 2008; Pehmer et al., 2015a; Resnick et al., 2010).

When studying effective teaching and learning processes and implementing changes in classrooms, especially in classroom discourse, individual student characteristics are highly relevant. Cognitive and motivational-affective student characteristics are strong predictors of student success in schools and learning in classrooms (Hattie, 2008). Pre-achievement is an example of an important cognitive characteristic, while the subject-related self-concept of ability is an example of a relevant motivational-affective characteristic. In particular, high pre-achievement and self-concept of ability are related to increased student engagement, the execution of deep learning activities, and positive experiences of motivation (Jurik, Gröschner, & Seidel, 2013; Lipowsky et al., 2009; Pauli & Lipowsky, 2007; Pehmer et al., 2015a; Turner, Warzon, & Christensen, 2011). Especially, student self-concept of ability is a key characteristic of students' beliefs about their own capabilities and affects their participation in classroom discourse (Jurik et al., 2013, 2014; Seidel & Shavelson, 2007). Students with a low self-concept of ability are less likely to be involved than students with a higher self-concept of ability (Jurik et al., 2013). In particular, it is hard for teachers to be aware of these low self-concept students and their needs because they have difficulties assessing student self-concept of ability and judging it accurately (Praetorius et al., 2013; Praetorius & Südkamp, 2017; Südkamp & Praetorius, 2017). This is particularly the case when teachers are mainly involved with students in tightly guided

teacher-student interactions with few possibilities for exploring individual student thinking (Pielmeier, Huber et al., 2018).

In this dissertation, two lines of research are linked. In Study I, adaptive functions in the relationship between accurate teacher judgment of student characteristics and observed verbal interactions between teachers and students were explored. It is assumed that adaptive teaching behavior will be found in the way that higher judgment accuracy is related to higher level of verbal interactions in classroom.

Furthermore, in a second study the constructs investigated first on a more general level, are analyzed in-depth in the second study. So far, none of the existing TPD programs targeting classroom discourse have attempted to understand how the individual changes in the classroom discourse quality of teachers is connected to an improved understanding of the learning characteristics of students. For this reason, in this dissertation in Study II, it was empirically explored with a case analyses how three low-performing teachers who participated in a video-based TPD program on productive classroom discourse undertook changes accordingly and the changes were then accompanied both by a positive development in the self-concept of ability of their students and an improvement in the teachers' accuracy when judging this student characteristic.

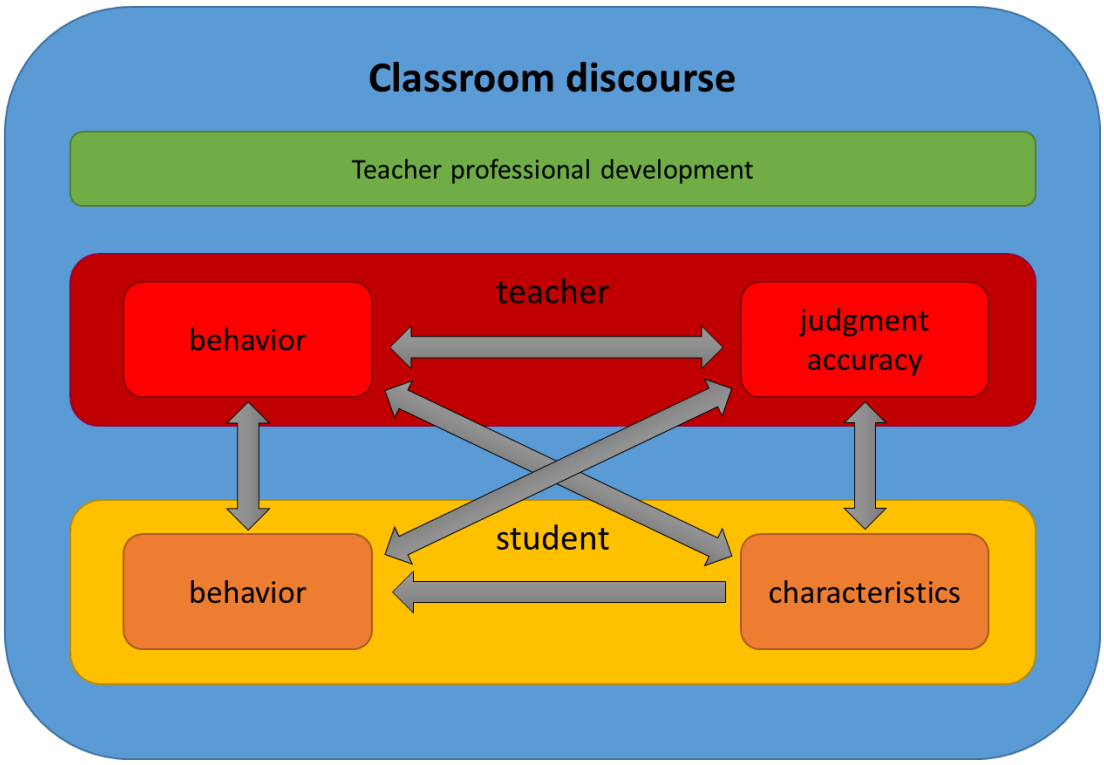

Figure 1. Framework model of this dissertation

Teacher (*light red box*) and student behavior (*orange box*) here in the context of classroom discourse (*blue box*) influence each other. On teacher side (*red box*), besides several other factors, their behavior is influenced by teacher judgments more specifically teacher judgment accuracies (*light red box*) and vice-versa. On student side (*yellow box*), these also influence student behavior (*orange box*) and student characteristics (*orange box*). Student characteristics affect teacher behavior and also student behavior. In the second study, teacher professional development program (*green box*) has an influence on the teacher, more specifically on teacher behavior.

1.2 Aims and Scope

Research has already brought important findings with regard to the elements illustrated in the framework model of this dissertation. Still there remain a number of open questions for research in each field. This dissertation connects different strands of research around teacher and student behavior in classroom discourse and focuses on exploring the effects of and on teacher judgment accuracies and student characteristics.

In the last decades, research in *classroom discourse* so far has mainly focused on describing teaching patterns (Howe & Abedin, 2013). This dissertation first clarifies the construct classroom discourse as there are many synonyms used in this field of research (Chapter 2.1). It is then further outlined how classroom discourse looks like and how it should look like (Chapter 2.2). Research on classroom discourse has shown that verbal teacher-student interactions as they occur during classroom discourse can have positive effects on student learning processes and developments (Chi, 2009; Mercer et al., 1999; Pehmer et al., 2015a; Turner, Meyer, Midgley, & Patrick, 2003; Webb et al., 2014). Within this context, especially *teachers questioning* behavior and *feedback* culture have shown to affect *student engagement* (Cazden, 2001; Chin, 2006; Pehmer, Gröschner, & Seidel, 2015b). Therefore, Study I addresses these constructs in detail. In Study II, the focus is more on the general quality of classroom discourse, more specifically on teacher classroom discourse performance. So far, classroom discourse has rarely been investigated under the perspective of classroom discourse as a learning opportunity for teachers to learn about their students (Chapter 2.3). One step towards closing this gap in research will be done with Study II.

Research on classroom discourse has revealed that there is room for improvement and that it needs to be improved (e.g. O'Connor et al., 2017) as it is the main teaching method in classrooms all over the world (e.g. O'Connor et al., 2017; Pehmer et al., 2015b). Therefore, several teacher professional development programs have evoken fostering classroom discourse (Pielmeier, Böheim et al., 2018). All these have in common that also teachers' understanding of student thinking was investigated and not only teacher knowledge and teacher practice changes (Borko, Jacobs, Eiteljorg, & Pittman, 2008; Pielmeier, Böheim et al., 2018; Santagata, 2009). Study II adds to this field of research as it is investigated how teachers change their classroom discourse practices through the TPD program. Furthermore, a student characteristic (self-concept

of ability) is analyzed as outcome and the development of teacher judgment accuracy of this characteristic.

Chapter 3.1 provides an overview of literature how *teacher professional development* is defined and what its core features are, leading to an overview of TPD programs in the context of classroom discourse (Chapter 3.2). The conceptualization and refinement of the Dialogic Video Cycle (DVC) – a video-based TPD program – applied in Study II is explained in Chapter 3.3. Study II investigates therefore, how three low-performing teachers implement the TPD content and how they have changed their classroom discourse performance after the DVC. Teacher professional development is a very sensitive and individual process (Borko, 2004; OECD, 2009), that is why teachers are investigated individually and then discussed in relation to each other at the end.

Classroom discourse is not only influenced by teacher behavior and after a TPD program by teachers' changes, but also student behavior and student characteristics play a central role. *Student characteristics* such as student *pre-achievement* as a cognitive characteristic and student *self-concept of ability* as a motivational-affective characteristic have been studied in the educational context as major pre-requisites for student learning success and process (Snow, Corno, & Jackson, 1996). The crucial role of these two mentioned student characteristics is reviewed in Chapter 4 with regard of each characteristic's role for classroom discourse. In the context of classroom discourse, student engagement has been shown to be strongly affected by student self-concept of ability (Huber, Häusler, Jurik, & Seidel, 2015; Jurik et al., 2013, 2014; Seidel, 2006). So far, often the relation of these two student characteristics and how they influence each other has been investigated (e.g. Marsh & Martin, 2011). It has not yet been investigated if and how a changed classroom discourse affects student self-concept of ability (Study II). As this is also a sensitive individual characteristic, the analysis are based on students entry level of self-concept of ability (low, medium or high) (for more details see Chapter 8). Previous studies have shown a strong relationship between student pre-achievement and student self-concept of ability (e.g. Marsh & Martin, 2011). Instead of investigating how these two characteristics are related together, this dissertation investigates the statistical interaction of each characteristic with teacher judgment accuracy on classroom discourse variables in Study I (see Chapter 7). Study II focuses only on self-concept of ability as this is seen and also has been proven in Study I as the most important characteristic when studying classroom discourse processes (Jurik et al., 2014).

Individualized instruction adapted to students' individual needs is widely requested (Corno, 2008). But therefore, teachers require diagnostic skills to be aware of their students' pre-requisites and to understand students learning and thinking processes (Corno, 2008; Machts et al., 2016). First, this dissertation clarifies the construct of *teacher diagnostic competences* (Chapter 5.1) and discusses the different ways this construct is conceptualized and measured. In a further step, teacher judgments and *teacher judgment accuracy* are outlined in Chapter 5.2 as one (most common) way to measure teacher diagnostic competencies. Finally, the relevance of accurate teacher judgments and their role for classroom discourse is elaborated (Chapter 5.3). Teacher judgments have been investigated in-depth since many years (e.g. Südkamp & Praetorius, 2017) because of its relevance for adjusting instruction (Praetorius & Südkamp, 2017; Schrader, 2017). This assumed relevance of teacher diagnostic competence can only be correct, if teachers accurately judge their students characteristics (Helmke, 2009; Praetorius & Südkamp, 2017).

So far, research on teacher diagnostic competences has not yet investigated if there is an adaptive teaching behavior with regard to classroom discourse and teacher judgment accuracy under control of the level of student characteristics. Furthermore, TPD in the context of classroom discourse has not yet investigated effects on or of teacher judgments. Furthermore, teacher judgment accuracy so far has mainly been represented as correlation of teacher judgment and student score of a certain characteristic. In this sense, teacher judgment accuracy on a general level is commonly calculated as correlation of teacher judgment and student characteristic data. In Study II, a new way to calculate teacher judgment accuracy on an individual level as difference of teacher judgment to student characteristic data was applied. This way was explored also in Study I, but due to similar results and to assure comparability with previous research, it was decided to stick with the common method. Whereas in Study II the individuality is more in focus, it made more sense to use also a measurement for teacher judgment accuracy on individual student level (based on Thiede et al., 2018).

Teacher judgment accuracy built together with classroom discourse the core of this dissertation. This dissertation aims to bring light in the complex interplay of classroom discourse and teacher judgment accuracy in relation to relevant student characteristics and on top with the impact of a video-based teacher professional development program. All aspects of the framework model of this dissertation, student characteristics, teacher judgments, as well to teacher and student behavior in classroom discourse and

additionally on teacher professional development add to the state of research in many small ways. The findings from the two empirical studies of this dissertation arise new questions for educational research and practice regarding teacher diagnostic competences and classroom discourse, but also for teacher professional development (see Chapter 9). Parts of the following chapters correspond to parts of the manuscripts that have been published or submitted for publication by the author of this dissertation in preparation of this dissertation (Pielmeier, Huber, & Seidel, 2018; Pielmeier, Seidel, Schindler, & Gröschner, submitted).

2 Verbal Teacher-Student Interactions – Classroom Discourse

In the 21st century, classroom discourse has gained attention in research as it is the main teaching method (Pehmer et al., 2015b). This chapter gives an overview of classroom discourse from students and teachers behavioral perspective.

In the framework model of this dissertation, classroom discourse is the overarching element within student and teacher behavior as core features of classroom discourse. Research on classroom discourse calls for teacher professional development programs to improve classroom discourse and to make it more productive. Classroom discourse also provides opportunities for teachers to learn not only with but also about their students. Within this dissertation two perspectives which can be used to analyze classroom discourse are brought together and each of their strengths are highlighted. This chapter defines the construct classroom discourse. It explains how productive classroom discourse should look like and how from that new opportunities appear for teachers to better get to know their students.

2.1 Classroom Discourse – Clarifying the Construct

Classroom discourse is the main teaching method in classrooms (Hiebert et al., 2003; Pehmer et al., 2015b). First of all, it has to be clarified what classroom discourse is and from which perspectives it can be looked at. Why is it called classroom discourse? There exist several phrases and expressions often used as synonyms for classroom discourse. *Classroom talk* or *verbal teacher student interactions in classroom* (Jurik et al., 2013; Mercer & Dawes, 2014) is used to define and specify that the talk takes place in classrooms. Classroom talk seems to be understood in a more neutral way of talking than *classroom dialogue* (Howe & Abedin, 2013) which expresses that learning is supposed to happen through talk. A *dialogue* indicates that at least two people are involved in a conversation about a certain topic (Howe & Abedin, 2013). *Dialogue* is

often used as synonym for conversation (Howe & Abedin, 2013). The opposite would be a *monologue* were only one person is talking and imposes his or her idea. Classroom dialogue would indicate students as equal conversational partners. Classroom discourse includes a variety of approaches and purposes. Therefore, several researchers argue for the distinction between *authoritarian* and *dialogic talk* (Aguiar, Mortimer, & Scott, 2010; Scott, Mortimer, & Aguiar, 2006). In this sense, classroom talk can be of certain quality with regard to equality of teachers and students. A further distinguishment of classroom talk is its regulative function implying a "teacher power or authority" versus an instructional function to construct knowledge (Buzzelli & Johnston, 2001). "The New London Group (1996) regards "discourse" as a sociocultural and political entity that subsumes ways of saying, writing, doing, being, valuing and believing. A discourse facilitates communication and establishes social identity within a community" (Bleicher, Tobin, & McRobbie, 2003).

In the context of research on teacher and student interaction the field shows many different expressions for classroom discourse. Based on the outlines above, within this dissertation classroom discourse is understood and defined broadly as the verbal teacher-student interactions as they occur in daily teaching practices during classroom instruction. It should be characterized by equal opportunities for students and teachers to interact. Either the teacher or a student can facilitate the discourse. A successful and productive classroom discourse "can enhance student understanding, add context to academic content, broaden student perspectives, highlight opposing viewpoints, reinforce knowledge, build confidence, and support community in learning" (Wikipedia, 2019). Research on classroom discourse has to take several points of view into consideration.

Classroom discourse itself has several facets, but also research on classroom discourse has different starting points. Like in this dissertation on the one side the state of the art – the way classroom discourse is – can be investigated and the other side its changeability – how classroom discourse can be changed through Teacher professional development (TPD).

2.2 Classroom Discourse – How does it look like?

Teaching and learning processes in classrooms are typically situated in contexts in which teachers verbally interact with students, and in this context most content instruction takes place (Mercer, 2008, 2010). Therefore, analyzing teacher and student behavior in verbal interactions is of particular interest. If implemented well, verbal teacher-student interactions have been shown to positively affect student learning processes and developments during the process of schooling (Chi, 2009; Mercer et al., 1999; Pehmer et al., 2015a; Resnick et al., 2010; Turner et al., 2003; Webb et al., 2014).

In many classrooms, however, teachers verbally interact with only a few students. Often these students have strong pre-requisites, such as high pre-achievement (Lipowsky, Rakocy, Pauli, Reusser, & Klieme, 2007; Pauli & Lipowsky, 2007). If teachers manage to equally distribute student engagement across diverse student groups in each class, more positive learning gains could be achieved (Lipowsky et al., 2007; Pauli & Lipowsky, 2007). Furthermore, students' conceptual understanding improves when many students are engaged in argumentative and interactive discourse settings (Chi, 2009; Mercer et al., 1999; Pehmer et al., 2015a, 2015a; Resnick et al., 2010; Webb et al., 2014). Thus, a main teaching goal in verbal interactions with students should be to involve and activate as many students in a classroom as possible (Walshaw & Anthony, 2008).

Many verbal interactions between teachers and students are initiated by teachers and are in the form of questions. Therefore, elaborating teacher questions that invite students to think deeply about a topic or a task are a particularly important teaching act (Cazden, 2001; Chin, 2006; Jurik et al., 2014; Pehmer et al., 2015a). These kinds of questions encourage students to reproduce known facts, link new information to existing knowledge and pre-experiences, and stimulate deep learning activities (Pehmer et al., 2015a; van Zee, Iwasky, Kurose, Simpson, & Wild, 2001). In addition, these questions also support students in the application of deep cognitive learning strategies (Alexander, 2005; Brown & Wragg, 2001; Lee & Kinzie, 2012; Pehmer et al., 2015a).

When teachers and students are verbally interacting, there are many events in which teachers provide explicit or implicit feedback. Therefore, teacher feedback is one of the most powerful and meaningful factors that can positively influence learning processes (Hattie, 2008; Hattie & Timperley, 2007; Jurik et al., 2014; Pehmer et al.,

2015a). Teacher feedback is characterized by providing constructive and supportive information during the learning process (Hattie & Timperley, 2007; Kobarg & Seidel, 2007; Timperley, 2013). In this way, feedback activates further student thinking and understanding (Gan & Hattie, 2014; Pehmer et al., 2015a) and can lead to positive development of motivational-affective student characteristics, such as self-concept of ability (Deci, Koestner, & Ryan, 1999; Jurik et al., 2014).

2.3 Teacher's Learning Opportunities during Classroom Discourse

In productive classroom discourse, there are multiple opportunities for teachers to get to know their students better. Thereby, the verbal activity of students in classroom discourse represents an important basis for teachers to be informed about their students (Borko, Roberts et al., 2008). Accordingly, teachers need to provide room for their students to elicit ideas, lines of argumentation, interests, and views about learning (Walshaw & Anthony, 2008). Student participation has been demonstrated to positively affect higher-order learning processes as well as learning outcomes as discourse can strengthen the conceptual understanding of a topic being discussed (Gomez Zaccarelli et al., 2018; Kuhn, 2010; Mercer, Dawes, Wegerif, & Sams, 2004).

Not only can students learn from productive classroom discourse, teachers can also learn about their students. Teachers learn from student responses and the responses that students give to each other. To use this advantage and create a network of student responses as actual discourse, classroom discourse has to meet a number of quality dimensions (Michaels & O'Connor, 2012). First, classroom discourse needs to be oriented towards a learning goal that students know (Alexander, 2005), and classroom discourse needs to be structured and purposeful (Seidel & Prenzel, 2006) so that students can contribute in a meaningful way (Resnick et al., 2010). Second, for productive classroom discourse, it is necessary for many students to be involved (Alexander, 2005; Gröschner, Schindler, Holzberger, Alles, & Seidel, 2018; Lipowsky et al., 2009; Pauli & Lipowsky, 2007; Walshaw & Anthony, 2008). Teachers need to activate students and open the classroom discourse by inviting many students to participate (Jurik et al., 2014; O'Connor et al., 2017; Oliveira, 2010; Pehmer et al., 2015a) and make them listen actively to each other. Third, it is the teacher's task to create an environment and setting

where students are not afraid to engage (Gomez Zaccarelli et al., 2018) and to link student contributions together (Osborne et al., 2016). Fourth, a supportive learning environment characterized by making room for and appreciating mistakes is required (Michaels & O'Connor, 2012), and this kind of learning environment involves feedback that provides students with information on their further learning processes (Hattie & Timperley, 2007). Teachers need to scaffold to further encourage students to actively participate in classroom discourse (Gomez Zaccarelli et al., 2018; Howe et al., 2007; Webb, 2009).

3 Teacher Professional Development to Improve Classroom Discourse

Teacher professional development is a highly requested professional demand. Recently, teacher preparation and teacher professional development point in the direction of structuring teacher education and teacher professional development around high-leverage practices (Ball & Forzani, 2009; Grossman, Hammerness, & McDonald, 2009; McDonald, Kazemi, & Kavanagh, 2013). The often mentioned gap between theory and practice, brought up more and more a shift from knowledge to practice especially in teacher education/preparation (Ball & Forzani, 2009). This can also be transferred to teacher professional development in the sense of life-long learning. Within this context often *high-leverage practice* is now a common idea. The term high leverage practices is defined "as a criterion for those practices most valuable for teachers to be able to learn and carry out" (Ball & Forzani, 2009). This means teachers should continuously learn in order to improve their teaching, in this context their classroom discourse.

Within this chapter it is explained how TPD programs should be conceptualized in order to make positive changes to classrooms. However, it is also explained how different professional development programs can change one of the most important high-leverage practice – classroom discourse. Ball and Forzani (2009) state that "Being able to teach well, […] depends on a flexible repertoire of high-leverage strategies and techniques that can be deployed with good judgment depending on the specific situation and context." Furthermore, the conceptualization and facilitation of a teacher professional development program – the Dialogic Video Cycle – is explained.

3.1 Teacher Professional Development Programs – Definition and Core Features

For changing classroom routines teacher professional development programs are indispensable (Kazemi & Hubbard, 2008). "Professional development is defined as activities that develop an individual's skills, knowledge, expertise and other characteristics as a teacher." (OECD, 2009). In order to achieve positive changes through a TPD program, the program has to fulfill several features. Desimone (2009) has identified five highly accepted core features of effective TPD programs: content focus, active teacher learning, coherence, duration, and collective participation.

- *Content focus* emphasizes the need for a defined goal for teacher learning.
- *Active learning* which includes opportunities for teachers to engage.
- *Coherence* points out the importance to take teachers' knowledge and beliefs into consideration.
- *Duration* means a sufficient amount of time including the span of time and the number of hours spent in the TPD program.
- *Collective participation* for a powerful way of teacher learning e.g. through interaction and discourse.

Continuity is also an important feature in order to assure, maintain and increase teachers' instructional quality (Vermunt & Endedijk, 2011; Vigerske, 2017). Furthermore, it is important that teachers see the relevance to their own classroom context (Putnam & Borko, 2000). The biggest challenge of teacher professional development programs is the bridge between the TPD program and teachers' daily routines. Video excerpts of teaching have been introduced as an innovative feature for effective TPD programs (Blomberg, Sherin, Renkl, Glogger, & Seidel, 2014; Borko, Jacobs et al., 2008; Seidel, Stürmer, Blomberg, Kobarg, & Schwindt, 2011; Sherin & Elizabeth van Es, 2002).

3.2 TPD Programs in the Context of Classroom Discourse

In typical tightly guided teacher-centered classroom discourse, students often give short answers or provide a keyword to the teacher (Seidel & Prenzel, 2006). These kinds of classroom discourse practices limit teachers to 'opening up windows' for teachers to learn about their students which requires that students actively participate. Then more accurate judgments of students' pre-requisites for learning can be made. The elements of dialogic teaching need to be implemented in the classroom discourse, in order to achieve more and more meaningful student contributions. Only then can teachers gain the opportunities to include more students in the discourse and to evaluate the quality of each contribution. Consequently, the learning processes and ideally students characteristics of students can be assessed and judged more accurately, resulting in adaptive, differentiated teaching (see Figure 2).

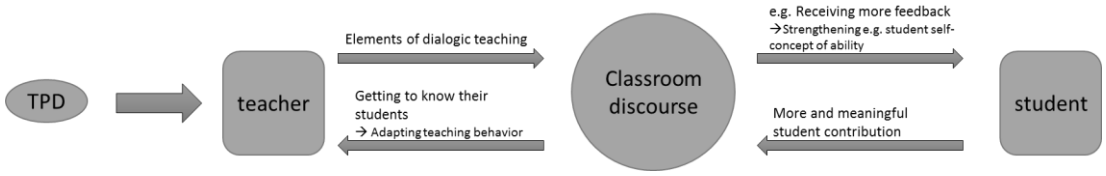

Figure 2. Assumed Mechanism (Pielmeier et al., submitted)

Several TPD programs for fostering classroom discourse, such as "CamTalk" (van de Pol et al., 2017), "Lesson Study Program" (Vrikki et al., 2017), "Accountable Talk" (Michaels et al., 2008), and the "ART" (Reznitskaya, Wilkinson, Oyler, Bourdage-Reninger, & Sykes, 2016), share in common the measurement of the program's effectiveness not only with an investigation of the changes in teacher knowledge and teaching practice but also in the changes in the teacher's understanding of student thinking (Borko, Jacobs et al., 2008; Santagata, 2009).

In addition, the above-mentioned programs all strive for the implementation of evidence-based standards for effective TPD. The meaning of the teachers' active learning in connection to their *individual* classroom routines (Putnam & Borko, 2000) is one of the biggest challenges for TPD programs (Desimone, 2009). Video excerpts of teaching have been introduced as an innovative feature for connecting new ideas to individual routines (Borko, Jacobs et al., 2008; Seidel et al., 2011; Sherin & Elizabeth van Es, 2002). Using video excerpts in TPD provides the opportunity to illustrate content in the context of classroom routines (Coles, 2013; Jacobs, Borko, & Koellner, 2009). Another

advantage of video is that it enables teachers to actively reflect on their own teaching practices or on those of their colleagues (Borko, Jacobs et al., 2008; Kleinknecht & Schneider, 2013; Seidel et al., 2011; Elizabeth van Es & Sherin, 2006). Video allows the capture of complex classroom situations in an authentic, practice-oriented manner for teacher learning (Gaudin & Chaliès, 2015; Elizabeth A. van Es, Tunney, Goldsmith, & Seago, 2014; Zhang, Lundeberg, Koehler, & Eberhardt, 2011) and closes the gap between practice and TPD content.

The careful facilitation of the discourse concerning video excerpts of the teaching practices of program participants is a core feature when using video as tool for TPD programs (Alles, Seidel, & Gröschner, 2018; Borko, Jacobs, Seago, & Mangram, 2014). In order to prevent the fear of appearing on video (Alles et al., 2018; Borko, Jacobs et al., 2008; Elizabeth A. van Es et al., 2014) and to make a video-based TPD program successful, it is the facilitators' task to create a positive learning atmosphere and a conversation culture (Borko et al., 2014; Gröschner, Seidel, Pehmer, & Kiemer, 2014).

3.3 The Dialogic Video Cycle (DVC)

The Dialogic Video Cycle (DVC) (see Figure 3) was invented as a TPD program which aimed to cover all so far known core features of an effective TPD program. The *content focus* of the DVC was classroom discourse, to improve teachers dialogic teaching practices and so also student engagement and learning outcomes. An *active learning* environment was given in every workshop as e.g. teachers brought their own ideas for the lesson, but also provided feedback to each other. *Coherence* was also given by taking teachers ideas, beliefs, knowledge and feelings into consideration when planning and reflecting the videotaped lesson. *Collective participation* was also given throughout the whole DVC. Teachers were monitored and moderated by a facilitator (Gröschner et al., 2014).

The DVC follows the heuristic of lesson planning – lesson teaching – lesson reflection inspired by the problem solving cycle (PSC) ((Koellner et al., 2007). This heuristic and the tool video sought to encourage teachers by providing insights into their individual classrooms and by developing a deeper understanding of how classroom discourse works for them and the other participating teachers. The obtaining of insights

into their own and their colleagues' classrooms was aimed at leading to a productive discussion regarding classroom discourse and student behavior. The observation of and reflection on the contributions and interactions of both individual students and an entire class were one goal of the TPD program. The *duration* of the program was a total of 22 hours, split in seven workshops throughout one whole school year, and thus was at the minimum range of what is described as effective in the TPD research (Gröschner, Seidel, Kiemer, & Pehmer, 2015; Lauer, Christopher, Firpo-Triplett, & Buchting, 2014).

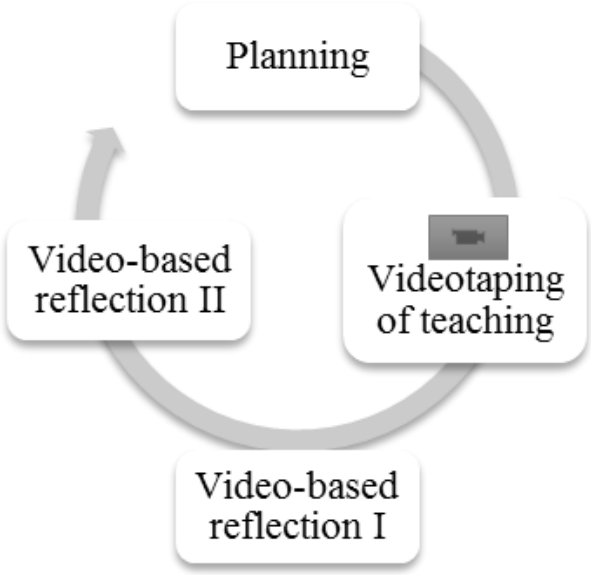

Figure 3. Dialogic Video Cycle (adapted from Gröschner et al., 2015)

4 Student Characteristics and their Role for Classroom Discourse

The most common way how teaching and learning takes place is through verbal interactions between teachers and their students in classroom discourse (Walshaw & Anthony, 2008). As also shown in the framework model of this dissertation, classroom discourse is thereby influenced not only by the teacher, but also by the students, students behavior, more specifically by students' characteristics as these define each student (*yellow box*). Each student is an individual and has different characteristic compositions (Huber et al., 2015). Researchers showed that student characteristics are important pre-requisites not only for student learning success and processes, but also for verbal teacher student interactions as they occur in daily classroom discourse. With regard to students' learning processes and learning success there are especially two characteristics, which have to be pointed out. These are explained in detail in the following two sub-chapters. Individual student characteristics, such as pre-achievement and self-concept of ability, are some of the strongest predictors of learning development and school achievement (Hattie, 2008) and have been studied extensively (Linnenbrink-Garcia, Pugh, Koskey, & Stewart, 2012; Seidel, 2006; Snow et al., 1996).

4.1 Student Pre-achievement as an Important Pre-requisite and Outcome of Classroom Discourse

During the course of schooling, students acquire subject-specific knowledge, and its acquisition is largely measured by content-related achievement (e.g. grades in subjects). In this sense, subject-specific knowledge can be regarded as an outcome of instruction. However, subject-specific knowledge also expresses a new entry point for future learning processes and is a pre-requisite for the next teaching acts in instruction (Hattie, 2008). Therefore, within this dissertation the term pre-achievement will be used.

Pre-achievement is a cognitive characteristic which has been shown to be an important characteristic for students learning success and students' performance (Dochy, Rijdt, & Dyck, 2002). In their study, prior knowledge explained 30% to 60% of variance in students' performance. Dochy and colleagues (2002) define prior knowledge as "the whole of a person's knowledge, which is as such dynamic in nature, is available before a certain learning task, is structured, can exist in multiple states (i.e. declarative, procedural and conditional knowledge), is both explicit and tacit in nature and contains conceptual and metacognitive knowledge components".

Students with high pre-achievement are found to be more frequently engaged in interactions with teachers (Jurik et al., 2013, 2014; Lipowsky et al., 2007; Seidel, 2006). Furthermore, students with high pre-achievement are more likely to verbally contribute to elaborating teacher questions (Ismail & Majeed, 2011). Regardless, high- and low-achieving students typically experience about the same amount of supportive teacher feedback (Hattie & Timperley, 2007). Lipowsky and colleagues (2007) showed that the higher students' pre-achievement was, the more learning success they achieved in class where students with average achievement level were underrepresented. Besides the effect of student pre-achievement on student behavior in classroom, it has also effects on students' performance (Dochy et al., 2002; Jurik et al., 2014). Students with high motivational-affective characteristics often also have higher cognitive abilities and higher pre-achievement. McElhone (2012) found that student talk enhanced learning and increased student achievement in text comprehension when students received varied and open-ended questions.

Students' pre-achievement is one individual student characteristic that is relevant when studying teaching and learning processes that occur during typical classroom instruction (Seidel & Shavelson, 2007). In order to facilitate learning, one fundamental principle teachers have to face, is understanding students' pre-achievement and adjust teaching on students' individual needs.

4.2 Student Self-concept of Ability as an Important Pre-requisite and Outcome of Classroom Discourse

Student self-concept of ability as another very important characteristic, especially for classroom discourse. Self-concept of ability is also highlighted as a particularly important student motivational-affective characteristic (Dai & Sternberg, 2004). As with pre-achievement, self-concept of ability serves as an important outcome of schooling as well as a relevant pre-requisite for further learning processes (Marsh & Martin, 2011; Marsh, Trautwein, Ludtke, Koller, & Baumert, 2006).

Student participation in classroom discourse has been shown to be strongly dependent upon the students' self-concept of ability (Seidel, 2006; Seidel & Shavelson, 2007). Self-concept of ability is formed by student's knowledge and perceptions in achievement situations (Dai & Sternberg, 2004). In fact, empirical studies have shown that students with a lower subject-specific self-concept of ability tend to be disengaged from classroom discourse despite having high pre-knowledge (Jurik et al., 2013; Pehmer et al., 2015b). One central problem occurring during classroom discourse is that students with low self-concept of ability are shy and do not respond to single answer questions as they exist in the typical IRF pattern of today's classroom discourse (Jurik et al., 2013). In contrast, it is typically the students with a higher self-concept of ability who are more frequently engaged in classroom discourse (Blöte, 1995; Jurik et al., 2013, 2014; Pauli & Lipowsky, 2007). Consequently, the opportunities that teachers have to learn about students are mainly focused on students who already have a positive view about themselves as learners. The problematic result is that teachers have very limited opportunities to learn about students with low self-concept of ability as these students rarely engage in interactions with their teachers.

Self-concept of ability has also been shown to be systematically related to specific aspects of verbal teacher-student interactions (e.g. (Jurik et al., 2013, 2014; Seidel, 2006)). Students more frequently engage in interactions with teachers when they have a higher self-concept of ability (Jurik et al., 2013, 2014; Pauli & Lipowsky, 2007). Furthermore, supportive teacher feedback is more often given to students with a high self-concept of ability (Blumenfeld, Pintrich, Meece, & Wessels, 1982; Ismail & Majeed, 2011) because these students participate more often. Students often receive evaluative

feedback (Cazden, 2001; Hattie & Timperley, 2007; Pehmer et al., 2015a), which can decrease students' self-concept of ability even more if they give an incorrect answer. In this sense, a so called 'Matthew'-effect often occurs, resulting in an increasing gap in further learning developments between students with higher or lower characteristics (Huber et al., 2015; Lipowsky et al., 2007; Pauli & Lipowsky, 2007). The 'Matthew'-effect originally describes the phenomenon that "the rich get richer and the poor get poorer" (Merton, 1968) and has been applied to the educational context in the way that "strong students get stronger and struggling students struggle even more" (Stanovich, 1986). Consequently, students with low self-concept of ability are involved even less, resulting in an increasing gap in the further development of learning between the students with a higher and those with a lower self-concept of ability (Huber et al., 2015; Lipowsky et al., 2007; Pauli & Lipowsky, 2007).

To avoid this development, it is important for teachers to be well aware of their students' characteristics, especially of their students' self-concept of ability because it is a strong predictor for the further development of learning, school success, and student participation (Marsh & Martin, 2011; Marsh, Trautwein, Ludtke, Koller, & Baumert, 2005). Therefore, it is also important to investigate how students' self-concept of ability changes within the context of TPD for more productive classroom discourse.

5 Teacher Judgments and their Role for Classroom Discourse

Adaptive teaching is described as teaching acts that are adjusted to the diverse needs of individual students (Corno, 2008; Dai & Sternberg, 2004). In order to adjust for diverse needs, teachers require diagnostic skills for continuously understanding the students' thinking processes in interaction with their characteristics (Corno, 2008; Machts et al., 2016; Praetorius et al., 2013; Seidel & Shavelson, 2007). Teacher diagnostic competences are in this context often conceptualized as teacher judgments more specifically teacher judgment accuracy (Schrader, 2017).

Referring to the framework model of this dissertation, teacher judgments (*light red box*) are part of teachers (*red box*) and influence teachers instruction, communication and behavior (Machts et al., 2016) and so classroom discourse. In the following chapter, it is outlined what teacher diagnostic competences are, how accurate teachers judge and the relationship between teacher judgments and verbal teacher-student interactions in classroom discourse.

5.1 Teacher Diagnostic Competences

Teacher diagnostic competences has gained attention in the last years because of its relevance for optimizing many pedagogic decisions, especially for students learning success and the effectivity of instruction (Schrader, 2017). Teacher diagnostic competence is a specific teacher competence that shapes the teaching processes and is required in order to create individual learning opportunities for students according to their individual needs based on their individual characteristics (Huber, 2017). Teacher diagnostic competence is defined as a cognitive achievement disposition which includes knowledge about conceptual, situational, procedural and strategic knowledge (Schrader, 2017).

The importance of teacher diagnostic competence is mostly justified with the goal of adaptive teaching (Praetorius, Greb, Lipowsky, & Gollwitzer, 2010; Praetorius & Südkamp, 2017). Instruction should be oriented towards students' individual needs based on their pre-requisites. For a continuous understanding of the interaction between students' thinking processes and their personal characteristics, teachers require diagnostic skills to adjust for the diverse needs of each student (Corno, 2008; Machts et al., 2016; Praetorius et al., 2013; Seidel & Shavelson, 2007). Research claims that successful teaching is only possible if adaptions to students' needs are made. Therefore, teachers have to diagnose these pre-requisites (Praetorius & Südkamp, 2017). Nevertheless, this theoretically assumed importance of diagnostic competence is only correct if teachers diagnoses are correct (Helmke, 2009; Praetorius & Südkamp, 2017).

5.2 Teacher Judgments and Teacher Judgment Accuracy of Student Characteristics

In empirical research, a tighter definition of teacher diagnostic competence is used, also due to practical reasons. Research on teacher diagnostic competences is mainly based on teacher judgment accuracy as measurement for teacher diagnostic competence (Praetorius & Südkamp, 2017; Schrader, 2017). The predictor and the criterion of the same characteristic is used to measure teacher judgment accuracy. It has become more common that teacher judgment accuracy is regarded as one manifestation of teacher diagnostic competence (Borko, Roberts et al., 2008; Kaiser, Retelsdorf, Südkamp, & Möller, 2013; Machts et al., 2016; Shavelson & Stern, 1981). In the empirical research, teacher judgment accuracy has been defined as the competence of a teacher to judge student or task characteristics accurately (Praetorius & Südkamp, 2017; Schrader, 2013; Spinath, 2005). This definition applies also to the research within this dissertation. Nevertheless, it is often criticized that teacher judgment accuracy and teacher diagnostic competence are used equivalent.

Not only with regard to measuring teacher diagnostic competence, there are also different approaches to measure and calculate teacher judgment accuracy (Spinath, 2005). Either a correlation between teacher judgment and student characteristic, or the average deviation is reported. Another component to measure teacher diagnostic competence is to measure teacher's diversity perception as a further element to capture

teachers awareness of several student characteristics (e.g. Huber, 2017; Huber & Seidel, 2018). Because of different measuring methods and discrepancies, it is difficult to speak about general judgment accuracy, it is better to talk of judgment accuracy of each single characteristic (Spinath, 2005). Within this dissertation there will be two ways applied to calculate teacher judgment accuracy (see more details in Chapter 6.2)

Accurate judgments of student characteristics are used to avoid the application of stereotypes and prejudices (Fiedler, Walther, Freytag, & Plessner, 2002). Earlier studies have shown that teachers displayed different levels of judgment accuracy when judging different types of student characteristics (Praetorius et al., 2013). Typically, teachers judge the level of student pre-achievement quite accurately (Demaray & Elliott, 1998b; Feinberg & Shapiro, 2003; Kaiser et al., 2013; Südkamp, Kaiser, & Möller, 2012). However, teachers show lower skills in accurately judging student self-concept of ability (Praetorius et al., 2013; Spinath, 2005; Südkamp & Praetorius, 2017). This difference in judgment accuracy might result from the kind of indicators that teachers refer to when judging individual student characteristics. Indicators for judging self-concept of ability might be less valid than those for judging pre-achievement (e.g., previous grades or test-scores as 'hard' facts for pre-achievement, whereas self-concept of ability is judged more through several cues which lead to a judgment (Brunswik, 1952; Koriat, 1997)).

Often student's performance overshines other personal characteristics (Huber et al., 2015) because it is easier for teachers to judge students on their pre-achievement, leading to a low- or high-performing categorization (Demaray & Elliott, 1998a; Feinberg & Shapiro, 2003; Kaiser et al., 2013; Südkamp et al., 2012). However, judging student motivational-affective characteristics, such as the subject specific self-concept of ability of a student, is more difficult for teachers (Praetorius et al., 2013; Spinath, 2005; Südkamp & Praetorius, 2017). These differences in accurately judging diverse student characteristics may result as already mentioned from the kind of indicators that teachers refer to when judging individual student characteristics (Pielmeier, Huber et al., 2018). As research has highlighted, student self-concept of ability is one particularly important factor among the motivational-affective characteristics of students (Dai & Sternberg, 2004), and so a correct judgment by teachers is a professional demand. In order to provide suitable supportive feedback to scaffold students, teachers have to be well aware of their students' self-concept of ability (Südkamp et al., 2012). Furthermore, teachers must also give more supportive feedback to students with a lower self-concept of ability in order to improve their views of themselves as learners.

5.3 Accurate Teacher Judgments of Certain Student Characteristics and their Role in Classroom Discourse

When studying verbal teacher-student interactions in classrooms, several theoretical considerations come into play in terms of integrating findings regarding teacher judgment accuracy and individual student characteristics. Accurate teacher judgments of student characteristics are important for knowing the level of pre-achievement and self-concept as aspects of class composition and can be regarded as an important professional competency for generally providing high teaching quality in a classroom (Shulman, 1987; Voss, Kunina-Habenicht, Hoehne, & Kunter, 2015), such as adaptive teaching behavior. Thus, accurate teacher judgments would lead to generally setting a high standard for verbal interactions with students. For example, specific questions can stimulate students' thinking and help them construct conceptual knowledge, which especially students with low pre-achievement need. Therefore, teachers need to accurately judge students' pre-achievement in order to provide frequent elaborating questions as one indicator for high quality in verbal teacher-student interactions (Chin, 2007; Kaiser et al., 2013; Ready & Wright, 2011). Also, the type of feedback and the way it is given can be effective in different ways (Hattie & Timperley, 2007). For providing supportive feedback to scaffold students, teachers also need to be well aware of their students characteristics (Südkamp et al., 2012). Thus, high teacher judgment accuracy can lead to adaptive teaching acts in the way that teachers intentionally pose more elaborating questions and provide more supportive feedback particularly to students with lower pre-achievement. In this sense, accurate judgments may lead to intentional decisions about a particular teaching script that is then applied to all students in a class.

Furthermore, accurate teacher judgments can help teachers to be aware of students with low self-concept of ability and purposefully provide those students with additional supportive feedback (Hattie & Timperley, 2007; M. Lotz & Lipowsky, 2015; Rakoczy, Klieme, Bürgermeister, & Harks, 2008). However, since previous research has shown that typically students with high self-concept of ability are the ones who receive additional supportive feedback compared to students with low self-concept of ability (Blumenfeld et al., 1982; Ismail & Majeed, 2011), accurate teacher judgments might simply lead to decreasing the gap typically found between those student groups. In this way, accurate teacher judgments might show an adaptive function with regard to the

support of particular student groups within the class context (Karst, Dotzel, & Dickhäuser, submitted, this issue).

Teacher expectations and judgments influence their discourse behavior (Machts et al., 2016; Sedova & Salamounova, 2016; Sedova, Sedlacek, & Svaricek, 2016). In the context of a TPD program on judging student thoughts, Alonzo and Kim (2018) showed that judgment quality and discussion quality are closely related. This might be a hint that teachers can only get to know their students better if the students are actively participating in a fearless atmosphere where every contribution is valued. Already in the 1970s, Brophy and Good (1970) have shown that teacher expectations affect teacher-student interactions as they occur in classroom discourse. Studies by Brühwiler (2017) and Sedova and Salamounova (2016) both showed that student participation is related to teacher judgment accuracy. This highlights the importance of investigating the teacher judgment accuracy of student self-concept of ability as this might also be an indicator of a teacher's learning process while getting to know students better with a changed teaching practice. Teachers need to be accurately aware of their students' characteristics as pre-requisites, and a change in teaching practice might also yield in an improvement in accurately judging student characteristics. This is why a teacher's accurate judgment of a student's pre-achievement and self-concept of ability might be especially relevant in teacher learning.

6 Research Agenda

6.1 Research Questions

This chapter gives a brief overview of the research question of this dissertation. The research questions were derived from previously presented research findings. This chapter provides an overview and a broad embedding of the research questions of this dissertation. The exact wording and the corresponding conjectures based on the presented theoretical background (Chapters 2 to 5) are presented for each study separately in the following Chapter 7 and Chapter 8.

The objective of this dissertation was to better understand the interplay of classroom discourse in its diverse manner and teacher judgment accuracy in relation to relevant student characteristics. First, the role of individual student characteristics in verbal teacher-student interactions was investigated in relation to teachers' judgment accuracy of relevant student characteristics to investigate the status quo of classroom discourse (Study I). Second, the effects of a video-based teacher professional development program on teacher's classroom discourse performance with regard to student self-concept of ability and the teacher judgment accuracy of this student characteristics of three teachers was investigated in-depth (Study II).

In reference to the framework model of this dissertation (see Figure 1), the following three research questions guided this dissertation:

1. How does classroom discourse look like and how can it be changed through a video-based TPD program?

2. How do student characteristics affect classroom discourse and how can student characteristics develop through a changed classroom discourse?

3. How do teacher judgment accuracies of selected student characteristics affect classroom discourse and what effects can a changed classroom discourse have on teacher judgment accuracy?

Student characteristics do not only influence students learning outcomes, but also students learning processes and students' behavior (Brühwiler, 2017). Especially two student characteristics have been highlighted as important when studying teaching and learning processes (Seidel & Shavelson, 2007). Students with high pre-achievement have been found to be more frequently involved in verbal interactions with teachers (Jurik et al., 2013, 2014; Lipowsky et al., 2007). Furthermore, previous research has shown that students with high self-concept of ability are more involved than students with low self-concept of ability (Jurik et al., 2013, 2014). Consequently, these students receive more feedback of their teachers, which is also known that this supports and increases students self-concept of ability (Blumenfeld et al., 1982; Ismail & Majeed, 2011). Based on these findings and assumptions, this dissertation investigated whether student pre-achievement and self-concept of ability and the judgment accuracy of these characteristics were predictive of verbal teacher-student interactions as they occur in classroom discourse within Study I. In this context, teacher judgment accuracies have been shown to be relevant for classroom discourse (e.g. Brühwiler, 2017). The focus of Study II was on the changeability of classroom discourse and how or if this also results in changes in teacher judgment accuracy. It is hard for teachers to change their routines (Duffy & Roebler, 1986). The effectivity and success of a video-based TPD program to make changes to the teacher's behavior and their classroom discourse performance is investigated within this study.

6.2 Methodological Approaches

Within this Chapter similarities and differences in the methodology of the two study of this this dissertation are outlined in order to create a general understanding of the results as preparation the general discussion in Chapter 9. The detailed methodology of each of the two studies are presented in Chapters 7 and 8.

6.2.1 Samples

Both student samples of the studies came from a high teaching track German Gymnasium (8^{th} grade for Study I, 9^{th} to 10^{th} grade for Study II). Students of this age were chosen because students of this age have already formed their characteristics, but they are still developing (Huber, 2017). This was particularly important as motivational-affective characteristics measured based on a student questionnaire with a self-report (more details are provided in the methods section of Study I and II in Chapters 7 and 8). Younger students have difficulties in expressing their motivational-affective characteristics in a reliable way, whereas older students can do so better (e.g. Marsh, Craven, & Debus, 1998). Furthermore, students' motivational-affective characteristics of this age group play a high role also with regard to students' career choices and their higher education of the following years (Wang, 2013).

German high-teaching track was chosen. German Gymnasium is said to be more homogeneous than other school, especially in other countries. Due to the fact o video-taping in classroom and the location of the researchers, Germany was chosen. Several studies and also the sample of Study I revealed a greater diversity especially with regard to motivational-affective student characteristics (e.g. Marsh et al., 2006).

Studying teacher judgment accuracy in this context is of particular interest as these high-level classrooms require fine-tuned diagnostic skills in order to estimate students individual learning pre-requisites and so to adjust instruction based on the students' needs.

6.2.2 Research Design

Study I and Study II were both settled within two larger research projects: Study I was settled within the project *Interaction*, funded by the German Research Foundation (DFG, SE 1397/7-1), Study II was nested in the project *Dialogue* II, which was also funded by the German Research Foundation (DFG, SE 1397/5-3). The *Interaction* study investigated the interaction of students' cognitive and motivational-affective characteristics and their impact on teacher-student interactions in the classroom.

The *Dialogue II* project is a TPD program-based research project on classroom discourse which examines how a video-based professional development program can change practices. Within this context, student characteristics, teachers' judgments and classroom discourse quality were assessed over one whole academic school year.

In both studies, students motivational-affective characteristics were assessed through a student questionnaire based on student self-report. Study I also included a cognitive student characteristic – student pre-achievement, but this was based on students' grade from the previous academic year, so it can be seen/regarded as a teacher judgment and not as an actual self-assessed student characteristic (for more details see Methods section of Study I).

Video data of Study I was coded very detailed, meaning each talking turn served as unit of analysis, whereas video data of Study II was coded on a more general level to assess classroom discourse quality in 10 minutes segments (for more details see Chapters 7.4.1 and 8.5.1).

Study I was based on a multi-method quantitative research design. Using different data sources from in total 20 classes allowed to investigate how initial student characteristics and its teacher judgment accuracy can affect or predict teachers' and students' behavior in classroom discourse in mathematics after a couple months into the school year.

Study II was designed as a longitudinal study and as qualitative case study. Case studies and small sample investigations have been established as important resources for generating a differentiated understanding of a teacher's individual uptake of a TPD program. O'Connor and colleagues (2017) showed in a comparison of two classes with different TPD conditions that there were no evident relationships between students with

or without vocal contributions regarding their effects on learning outcomes (student test scores), thus introducing the idea that active listening is another important aspect of student engagement in classroom discourse. In another case study, Gomez et al. (2018) described in detail how one teacher successfully implemented specific strategies from a TPD program, leading to productive classroom discourse. However, a study by van den Bergh et al. (2015) illustrated how some teachers have struggled with directly applying learning and self-regulation activities. One teacher showed directed and another undirected learning patterns, but both were still able to demonstrate the positive effects of the TPD program on their classroom behavior in the end. These results illustrate how case studies can provide deeper insights into individual teacher development and the implementation of TPD program contents, serving as a solid basis for further research.

6.2.3 Instruments

In both studies one lesson units of 45 minutes were videotaped for each measuring point. In Study I an event-sampling based video manual was used, where teacher or student talking turns served as the unit of analysis. In Study II, due to economic reasons and by following the current approaches in the field of classroom research (Reznitskaya, Wilkinson, Oyler et al., 2016), the time sampling method was used, where each lesson was split into four 10-minute segments which served as unit of analysis. Each study had its own video manual which was applied to each unit of analysis.

Both studies applied similar instruments to measure student characteristics. For Study I, a questionnaire four-point Likert scale from PISA (Programme for International Student Assessment) was used to determine student self-concept of mathematical ability (Hertel, Hochweber, Mildner, Steinert, & Jude, 2014). Study II applied a student questionnaire from Marsh et al. (2005) on a four-point Likert scale for assessing students subject-specific self-concept of ability. As cognitive characteristic in Study I, student pre-achievement was considered in terms of the students grades from the previous academic year. Pre-achievement assessing through grades is discussed critically in the field of research as pre-achievement is a high predictor for students learning success (Hattie, 2008; Marsh & Martin, 2011) and grades are already based on teachers' judgments. Therefore, only student self-concept of ability was considered in Study II as this is one of the most important student characteristics with regard to student behavior in classroom discourse (e.g. Jurik et al., 2014).

Teacher judgments were assessed through questionnaires. In Study I, teachers had to judge their students on a three-point Likert scale (details on gradation are provided in section 7.4.3). Study II was in this sense modified that teachers had to judge each individual student on a five-point Likert scale. In both studies, teachers were provided with definitions of each characteristic (details on gradation are provided in section 8.5.3). In Study I, teacher judgment accuracy was calculated as correlation of teacher judgment and student characteristic. This is the most commonly applied method (Machts et al., 2016; Praetorius et al., 2011; Praetorius et al., 2013; Spinath, 2005; Südkamp et al., 2012). In order to interpret and compare results better with previous findings from other researchers, this approach was applied. Nevertheless, it was also examined how a individual measurement/calculation of teacher judgment accuracy can be applied as there is much movement in research on teacher diagnostic competence and also lots of critique with regard to the way teacher judgment accuracy is assessed. As Study II focuses more on teachers' individual changes and also on student subgroups within each class, it was therefore decided to apply the calculation of teacher judgment accuracy as mean for each subgroup of the differences between teacher judgment and student self-report for each student. The calculation the teacher judgment accuracy is explained in detail in the Methods section of each Paper (see Chapters 7.4.3 and 8.5.3). At this point of the dissertation, it should provide a solid picture of the differences and serves as basis for the general discussion (Chapter 9).

6.2.4 Statistics

For both studies, data from student and teacher questionnaire were transformed into an suitable scale level for the later applied analyses. Methods used in this dissertation include, multi-level random intercept models, correlations, frequency analysis and interclass correlations. Therefore, R (R Foundation for Statistical Computing, 2016), SPSS (Version 24 und 25) (IBM Corporation, 2016, 2018), and Microsoft Excel (2017 und 2019) were used. Details on the models and software are given in the Methods sections of Chapters 7 and 8.

Statistics in Study I were on an advanced quantitative level as there were in total 18 classes with their 18 teachers and 501 students. Whereas statistics in Study II were on a basic statistical level with mainly descriptives as it was conducted as a case study analysis of three selected cases.

6.3 Relevant Publications

The empirical studies of this dissertation (Study I and Study II) have been published or submitted for publication in the following way:

Study I: The Influence of Teacher Judgment Accuracy on Verbal Teacher-Student Interactions in Classrooms

The effects of teacher judgment accuracy in interplay with student characteristics on verbal teacher-student interactions was investigated in the context of Study I. The manuscript was submitted to *Teaching and Teacher Education* and after revision accepted for publication on January 16th 2018.

Pielmeier, M., Huber, S., & Seidel, T. (2018). Is teacher judgment accuracy of students' characteristics beneficial for verbal teacher-student interactions in classroom? *Teaching and Teacher Education*. (76), 255–266. https://doi.org/10.1016/j.tate.2018.01.002.

Conception, preparation, analysis, and the publication-based presentation of Study I were fulfilled in the context of this dissertation (60%). The originating process, the preparation, and the presentation of Study I were advised by both co-authors (Sina Huber 20%; Tina Seidel 20%).

Study II: The Changeability of Classroom Discourse (through TPD) and its Effects on Teachers Judgments and Student Self-concept of Ability

Changes in teacher's classroom discourse performance and its effects on students' self-concept of ability and teacher's judgment accuracy of this student characteristic were in-depth analyzed in the context of Essay 2. The manuscript was submitted for publication to *Teaching and Teacher Education* on March 11th 2019 and is under review since April 16th 2019.

Pielmeier, M., Seidel, S., Schindler, A.-K., & Gröschner, A. (submitted). Opening 'Windows' for Teachers to Change Classroom Discourse. *Teaching and Teacher Education*.

Conception, preparation, analysis, and the publication-based presentation of Study II were fulfilled in the context of this dissertation (70%). This process was supported and advised by the co-authors Tina Seidel (15%), Ann-Kathrin Schindler (10%) and Alexander Gröschner (5%).

7 Study I: The Influence of Teacher Judgment Accuracy on Verbal Teacher-Student Interactions in Classrooms

7.1 Addressed Research Questions and Conjectures

This study investigated to what extent accurate teacher judgment of student pre-achievement and self-concept of ability are systematically linked to verbal teacher-student interactions in classrooms, including teaching acts such as verbal student engagement, elaborating teacher questions, and supportive teacher feedback. With this focus, the study explores the interplay of individual student characteristics, accurate teacher judgment of these characteristics, and verbal teacher-student interactions. In particular, the following two research questions were addressed:

1. Are individual student characteristics such as pre-achievement and self-concept of ability predictive of verbal teacher-student interactions in classrooms?
2. Are accurate teacher judgments of these two characteristics predictive of verbal teacher-student interactions? In particular, is the interplay between student characteristics and teacher judgment accuracy predictive with regard to verbal interactions, indicating adaptive functions of teacher judgment accuracy?

1) It is hypothesized that individual student characteristics such as pre-achievement and self-concept of ability positively affect verbal student engagement in interactions with teachers. In addition, these student characteristics are also expected to lead to more frequent experiences of elaborating teacher questions and supportive teacher feedback.

2) It is expected that high teacher judgment accuracy of pre-achievement and teacher judgment accuracy of self-concept of ability are positively related to verbal teacher-student interactions, indicating that teachers with high diagnostic skills generally facilitate on a higher level of verbal interaction in the way of adaptive teaching. With regard to pre-achievement, it is hypothesized that teachers with higher judgment accuracy provide more elaborating questions and supportive feedback particularly to low-achieving students. Furthermore, teachers with higher judgment accuracy of self-

concept of ability are expected to provide more supportive feedback for students with low self-concept of ability, indicating adaptivity with regard to students' need for additional support. For predicting such adaptive teaching behavior, teacher judgment accuracy in interplay with the level of student characteristics present in a classroom need to be studied.

7.2 Sample

This study was conducted with eighth-grade students (N_S = 501) and their mathematics teachers (N_T = 18) in the classrooms (N_C = 18) of a high teaching track German Gymnasium. Since the sample participated in a video study, there were some cases of missing video data and student characteristics data (e.g. students not being present at one of the measuring points due to illness etc.), the final sample consists of 348 students with full data sets of all measuring points. The mean student age was 13.79 years (SD = 0.51). In the German school system, students are grouped into different school forms (high, intermediate, low) after grade 4 to provide schooling for more homogeneous student compositions. Despite this selective process, a number of studies have shown still quite strong student diversity with regard to pre-achievement and self-concept of ability, indicating a large overlap in student compositions between school forms (e.g. between high and intermediate levels) (Hammer et al., 2016; Reiss, Sälzer, Schiepe-Tiska, Klieme, & Köller, 2016; Sälzer, Reiss, Schiepe-Tiska, Prenzel, & Heinze, 2013; Schiepe-Tiska & Schmidtner, 2013). This requires teachers to continuously monitor student characteristics in their classrooms even though classes are supposed to be more homogeneous compared to other school systems. These circumstances are of particular interest for studying teacher judgment accuracy in the process of teaching since these high-level classrooms require fine-tuned teacher diagnostic skills.

Participation in the study was voluntary. Half of the sample was selected randomly and was invited to participate in the study. The other half was recruited from a school network of the local university where university-level students of teacher education were assigned to internships. The students and teachers in the sample were representative of high-level German Gymnasium classes in terms of age, gender, family background of the children, and characteristics of participating teachers (mean age,

teaching experience, gender) (Jurik, Häusler, Stubben, & Seidel, 2015). All classrooms were located in urban and suburban areas of southern Germany. The 18 participating teachers had a mean age of 39.83 years (SD = 9.97). Their average teaching experience was 10.58 years (SD = 8.21).

7.3 Research Design

Data for this study were collected during the 2013/2014 academic year (see Figure 4). The study used a multi-method research design, which included measurements of both teachers and students as well as behavioral data from videos of instruction. At the beginning of the school year, student pre-achievement and self-concept of ability were assessed through a student questionnaire. After two months of instruction and after teachers became familiar with their students, teachers completed a questionnaire judging pre-achievement and self-concept of ability for each student in their classroom. Following these assessments, one lesson unit on the mathematical topic of "functional relationships" (e.g., proportionality) or "intercept theorems and similarity" was videotaped in each classroom. The topics are typically part of the national curriculum for eighth grade classrooms (Kultusministerkonferenz, 2003; Staatsministerium für Unterricht und Kultus, 2004). Videotaping occurred three to four months after the start of the school year. These 45-minute lesson units were chosen to provide a comparative learning environment across all classrooms and to observe teacher and student behavior in verbal interactions. On average 83.7% of the lesson were teacher-led whole group discussions. The lessons were typical (Kunter et al., 2006; Seidel, 2014; Seidel, Rimmele, Rolf, Prenzel, & Manfred, 2003) as there were also small-group activities (0.5%) and silent work (15.7%) included, but the classroom discourse was the predominant teaching method.

Using this research design, it was possible to investigate to what extent initial student characteristics and teacher judgment accuracy at the beginning of a school year can predict verbal interaction behavior in a mathematics lesson unit a couple of months into the school year.

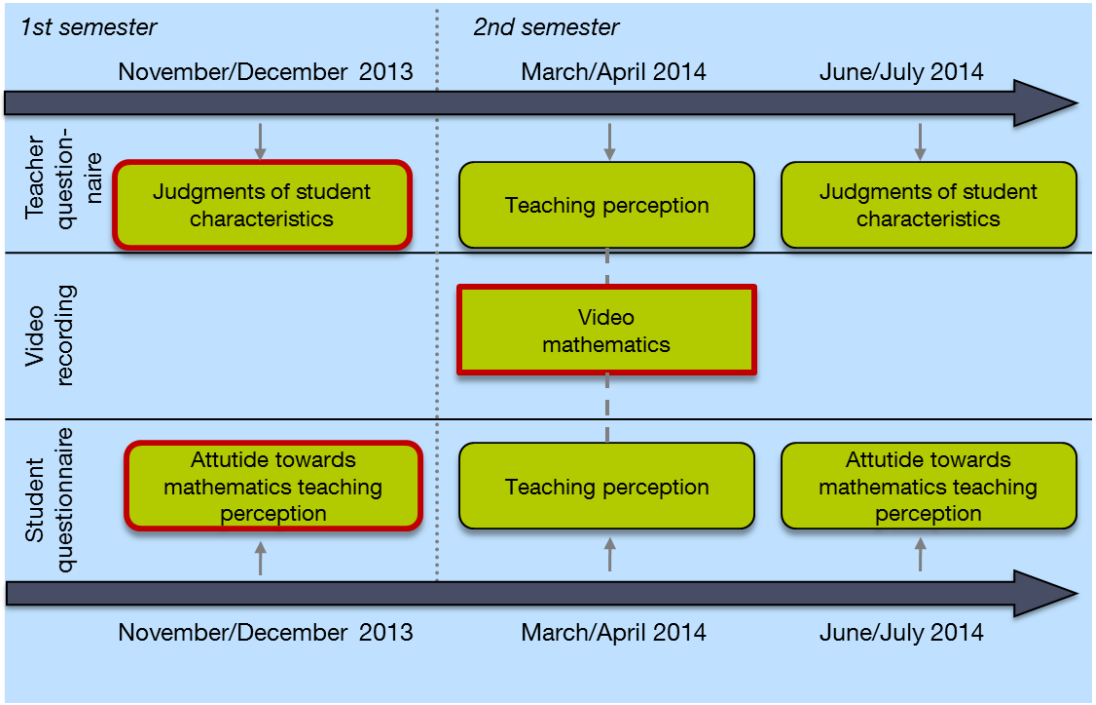

Figure 4. Study Design and Data Sources Study I (adapted from Jurik et al., 2015)

7.4 Instruments

7.4.1 Verbal Teacher-Student Interactions

Verbal teacher-student interactions were analyzed using systematic video analysis (Seidel, Prenzel, & Kobarg, 2005) and the software Interact (Mangold, 2014). Video analyses were conducted by two systematically trained, independent coders who were tested for reliable observations. Teacher and student talking turns served as a unit of analysis. These units where further coded according to a detailed coding manual to student engagement, elaborating teacher questions, and supportive teacher feedback.

Student engagement was coded as each verbal student contribution and was also represented as an analysis unit of student talking turns. This code was used to identify any content-related student statement (e.g., “I don’t think that this is the correct result, because mine is different”). Additionally, each student in a class had an individual code. This code was used when a student was verbally engaged in order to match verbal engagement and teacher interaction with other student-related data such as individual

characteristics. Of the student contributions, 95% could be assigned exactly to an individual student code. Reliability between coders for student engagement was 87.7% (Cohen's Kappa = 0.82).

Elaborating teacher questions were coded as a form of teacher question that initiated an interaction with one student or more. These questions included ones that allowed students to connect different ideas and concepts and were challenging enough to trigger further knowledge building (Oliveira, 2010; Pehmer et al., 2015a). The coding instructions defined elaborating questions as follows: "Elaborating questions are questions that intend to explain or elaborate facts that are not yet known in this form. The answers to these questions require deeper thinking." An example of elaborating questions is: "How would you do that if you did not have it [the graph] on the computer, but for example printed out small and you want to enlarge it? Does anyone have an idea how to do it?" Reliability between coders for this category was 79.5% (Cohen's Kappa = 0.76).

Supportive teacher feedback was coded when teachers responded to student engagement in a supportive way. Because feedback was generally rather scarce in these mathematics classrooms (as also found in other studies, cf. Klieme et al., 2008; Lipowsky et al., 2009; Voerman, Meijer, Korthagen, & Simons, 2012), diverse forms of feedback (Hattie & Timperley, 2007) were not differentiated further. For this study, feedback could involve either simple, positive confirmation or more elaborated positive and constructive feedback. Supportive feedback was defined as either "a short contribution by the teacher to a student's articulation" (e.g., "Yes" or "That's right") or "a response that gives the student a clue as to what was right in his or her statement and how to 'get to the learning goal' even better." In addition to confirmation of a student's statement, feedback also included support that could be both content-related and process-related. These forms of feedback gave the student the opportunity to review and improve his or her approach or response. An example of this was: "Yes, that's right, you explained it very well and that's correct." The amount of observed supportive feedback was used as an indicator of supportive teacher feedback. The reliability between coders was 81.7% (Cohen's Kappa = 0.66).

7.4.2 Student Characteristics

Student pre-achievement: Mathematics grades at the end of the previous school year were used as an indicator of student pre-achievement. Grades in Germany range from 1 (best) to 6 (worst). In this study the grades were recoded using a scale from 0 to 5, with low values indicating low pre-achievement and high values indicating high pre-achievement. Using this recoded scale, mean pre-achievement was 2.77 (SD = 0.90).

Student self-concept of ability in mathematics: A student questionnaire scale from PISA (Programme for International Student Assessment) was used to determine student self-concept of mathematical ability (Hertel et al., 2014). The scale included four items rated on a four-point Likert Scale ($\alpha = .92$, M = 2.39, SD = 0.81). An example item was: "In mathematics, I learn quickly" (Range: 1 = strongly disagree to 4 = strongly agree).

7.4.3 Teacher Judgment Accuracy

Teachers rated each individual student in their mathematics class on their level of pre-achievement and self-concept of ability in mathematics on a three-point scale (1 = low, 2 = intermediate, 3 = high). This scale was constructed based on an exploratory pilot study that indicated that teachers typically grouped their students according to these levels. The following instructions were given to the teachers: "Please assess pre-achievement and self-concept of ability of your students with regard to mathematics on a scale from 1 to 3 (1 = low, 2 = intermediate, 3 = high). To do this, enter your rating from 1 to 3 in the marked field for each individual student characteristic." Thus, teacher ratings were based on the teachers' own perceptions of students in their classrooms. On average, teachers rated their students as having intermediate pre-achievement (M = 1.95, SD = 0.69) and self-concept of ability (M = 2.02, SD = 0.66). To summarize, teacher data regarding students' pre-achievement and self-concept of ability refers to three groups: top quartile = 3 = high, bottom quartile = 1 = low (as outlined above). To match the distribution of teacher judgments with student data, students' grades as a measurement of pre-achievement and student self-report of self-concept of ability were also grouped into three percentile groups (bottom quartile = 1 = low, 2 intermediate quartiles = 2 = intermediate, top quartile = 3 = high).

For calculating teacher judgment accuracy, explored two different methodological ways were explored. First, the correlation (Spearman) of teacher judgment and student self-report within each class was calculated for each of the two characteristics (pre-achievement and self-concept of ability). This method is currently a quite common way to measure teacher judgment accuracy (Machts et al., 2016; Südkamp et al., 2012) and has been applied in several studies with similar study designs (Praetorius et al., 2011; Praetorius et al., 2013; Spinath, 2005). Nevertheless, other methodological approaches are currently also under investigation (cf. Südkamp, Praetorius, & Spinath, 2017) that can provide additional information, for example, on the individual student level. Therefore, a second methodological approach on the individual student level was also applied for this data set. In this approach, the difference between teacher judgment and student data was calculated on the level of each individual student and used as an indicator for teacher judgment accuracy. In comparing both methodological approaches, however, no systematic differences, with regard to model fits, explained variances and patterns of findings were observed. In this paper we, therefore, decided to present findings based on the first methodological approach since findings can be better compared to previous research in this field (Praetorius et al., 2011; Praetorius et al., 2013; Spinath, 2005). However, in discussing the findings at the end of the paper, the need for further research addressing multiple indicators for teacher judgment accuracy is also acknowledged.

7.5 Data Analysis

Basic statistical analyses were conducted using SPSS version 24 (IBM Corporation, 2016). Further statistical analyses were performed using R version 3.3.2 (R Foundation for Statistical Computing, 2016).

Descriptive Analyses. For descriptive statistics, means and standard deviations of student characteristics, teacher judgments, and verbal teacher-student interactions were calculated. To estimate intra-class correlations (ICC), the R package *ICC* with the function *ICCest* was used. Intercorrelations between the investigated variables were reported using the Spearman correlation coefficient (using SPSS).

Multi-level Analyses. Given the hierarchical structure of the data, multilevel analyses were applied to limit possible distortion from the dependency of the observations (Bickel, 2007; Hox & Kreft, 1994). In addition, the variance of independent variables between analysis levels was calculated. For the random intercept multilevel analyses, the R package *nlme* with the function *lme* was used. To calculate the marginal and conditional R^2, the R package *piecewiseSEM* and the function *sem.model.fits* was used (Nakagawa & Schielzeth, 2013). All data were standardized in R using the function *scale* from package *base* with the function *cbind* from package *base.* As outlined above, in total 153 missings were excluded from analysis. Missing values either referred to missing individual student characteristic data (N= 71), teacher judgment data (N = 37) or student behavioral data (N = 64).

For each of the three dependent variables (verbal student engagement, elaborating teacher questions, supportive teacher feedback), three multi-level models were tested. In Model 1, the two student characteristics were included (on student level) for their effect on behavioral measures. In Model 2, teacher judgment accuracy of individual student characteristics was covered (on class level). In Model 3, student characteristics (on student level), teacher judgment accuracy (on class level), and interactions among these indicators were included. In the following, it is spoken about the interplay between these indicators in order to avoid misunderstanding with regard to verbal teacher-student interactions as another important concept of this study. From a statistical point of view, these refer to interaction effects.

The marginal explained variance (R^2) was 0.02 for all dependent variables in Model 1. For Model 2, it ranged between 0.00 (for elaborating teacher questions and supportive teacher feedback) and 0.02 (for student engagement). In Model 3, marginal R^2 was 0.04 for elaborating teacher questions and student engagement, but for supportive teacher feedback, R^2 was 0.03. This means that up to 4% of the total variance of the models can be explained by the fixed effects of each model.

Even though the amount of explained variances is very low, empirical researchers argue that small R^2 values are expected when using behavioral data for empirical analyses (Blut, 2008; Cohen, 2010; Greene, 2012; Michaelis, 2009). The conditional R^2 is the total variance of the models explained by the total variance of the fixed effects and the sum of random variance components for each level of the random factor. The conditional R^2 values increased throughout the entire model. In the conditional R^2, the

variance explained by the class level was added. Conditional R^2 values ranged from 0.11 to 0.38. This means that 11% to 38% of the total variance of the models can be explained by both fixed and random factors. However, explained variances in this study are overall rather low and have to be kept in mind when interpreting the findings.

7.6 Results

7.6.1 Descriptive Analysis

7.6.1.1 Verbal Teacher-Student Interactions

The mean number of statements indicating verbal student engagement was 72.94 (SD = 34.28) statements in a lesson unit of 45 minutes. This was measured as the number of all content-related student statements in a class. Each individual student made an average of 3.77 (SD = 3.94) content-related statements per lesson. On average, 23.72 (SD = 17.54) elaborating teacher questions were asked per lesson. The majority of these questions were posed to the whole class and only a few were directed to an individual student (mean = 1.22, SD = 1.88). During the videotaped lesson, students received on average 1.47 (SD = 1.64) instances of directly addressed supportive teacher feedback per student with an average of 28.44 (SD = 15.08) total instances of supportive feedback per class.

Regarding an average individual student, student engagement (3.77 statements per student), individually experienced teacher initiating behavior of questioning (1.22 per student) and teacher responses in the form of supportive feedback (1.47 per student) were scarce events. In this study, it was investigated to what extent these events can be predicted by individual student characteristics as pre-requisites and teacher judgment accuracy as diagnostic skills.

Figure 5 provides an illustration of the relative frequency of verbal teacher-student interactions during the observed lesson, showing a fairly equal distribution of these events over time. The relative frequencies refer to the proportion of observed events in relation to the overall observation category system: Student engagement in relation to overall engagement between teachers and students, teacher elaboration questions in

relation to multiple kinds of teacher questions, and supportive teacher feedback in relation to other kinds of feedback. In terms of variation among classrooms in this sample, ICC values ranged from 0.10 for student engagement, 0.23 for elaborating teacher questions, and 0.19 for supportive teacher feedback. These variations are shown in Figure 6, which illustrates the amount of each teacher and student behavior in verbal interactions in each class.

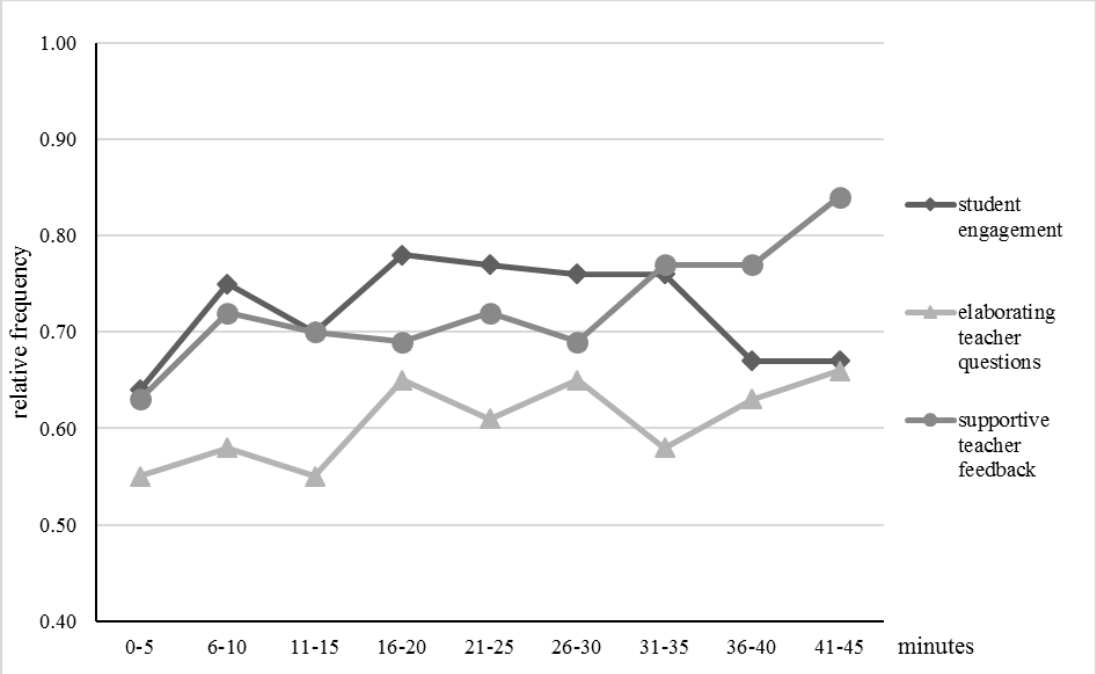

Figure 5. Distribution of relative frequencies of verbal teacher-student interactions during the videotaped lesson in 5 minute units (Pielmeier, Huber et al., 2018)

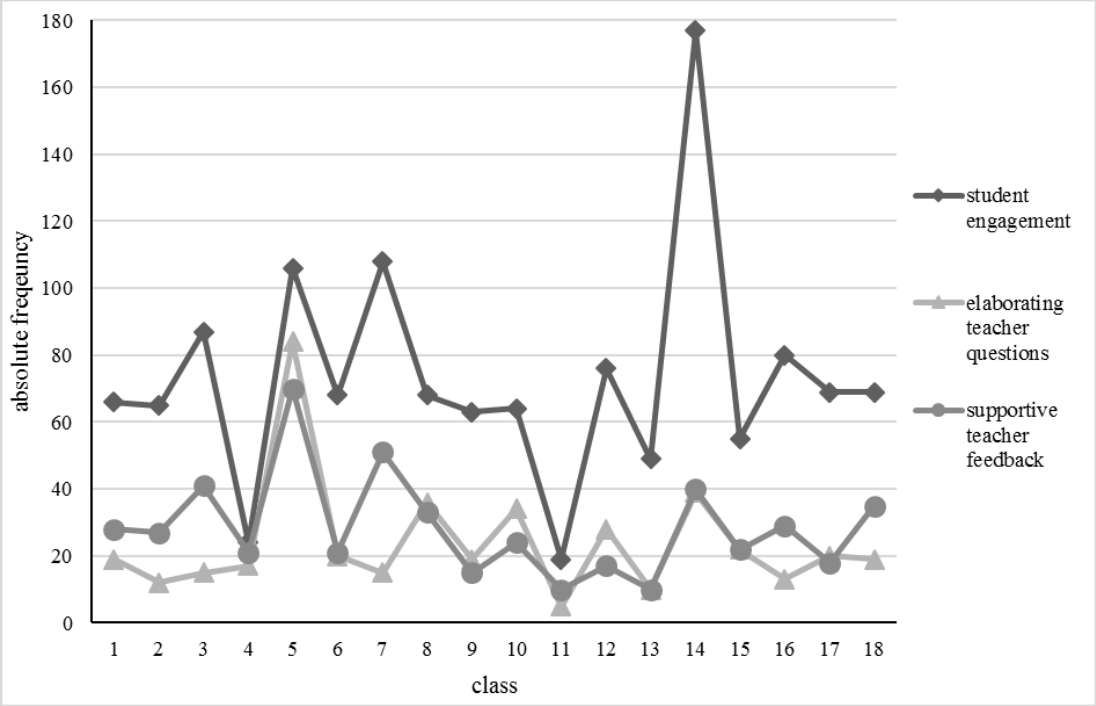

Figure 6. Distribution of absolute frequencies of verbal teacher-student interactions in each observed classroom (Pielmeier, Huber et al., 2018)

7.6.1.2 Student Characteristics

Figure 7 shows the distribution of student pre-achievement and Figure 8 illustrates the distribution of student self-concept of ability in this sample. Pre-achievement has a mean of 2.91 (SD = 1.03) and self-concept of ability a mean of 2.46 (SD = 0.84). These mean values and distributions are also present in each of the 18 individual classrooms.

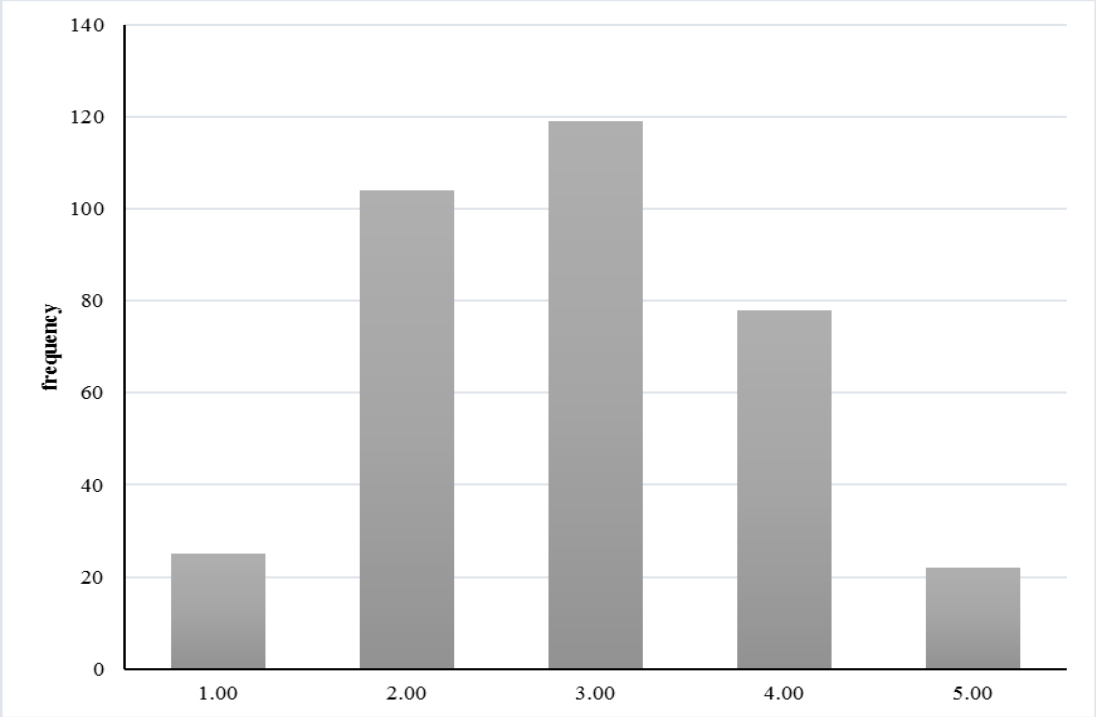

Figure 7. Distribution of absolute frequency of student pre-achievement on student level (1 = lowest grade, 5 = highest grade) (Pielmeier, Huber et al., 2018)

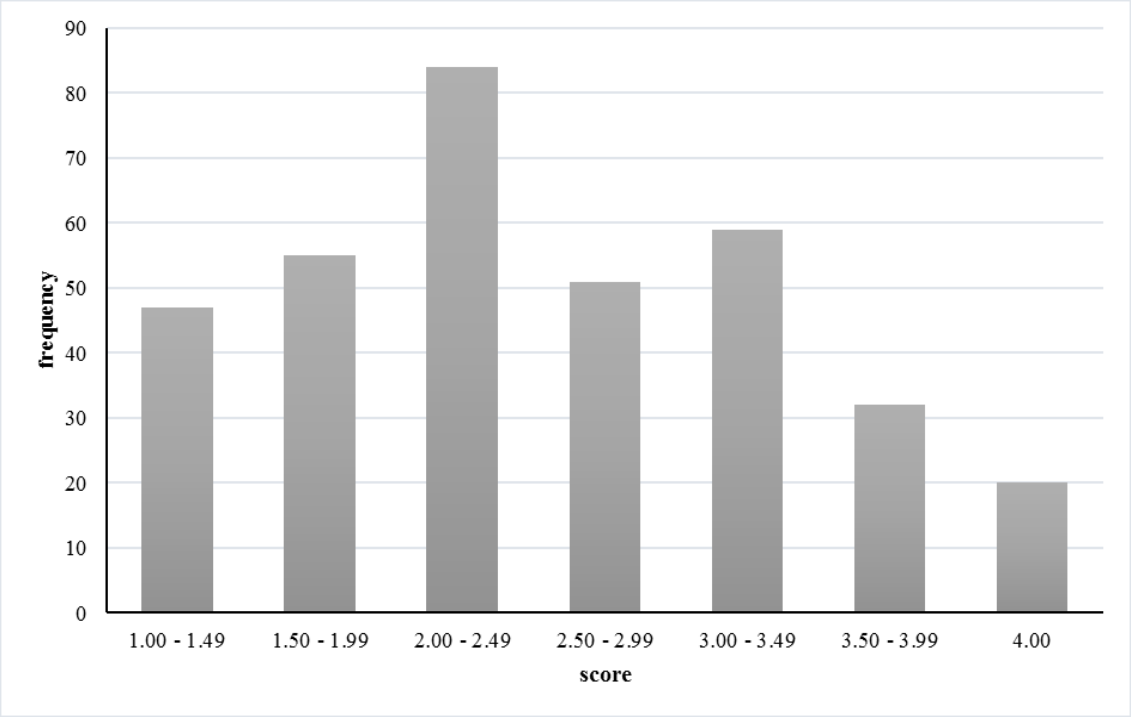

Figure 8. Distribution of absolute frequency of student self-concept of ability on student level (1 = low self-concept of ability, 4 = high self-concept of ability) (Pielmeier, Huber et al., 2018)

7.6.1.3 Teacher Judgment Accuracy of Individual Characteristics

Figure 9 shows the distribution of teacher judgment accuracy scores across the 18 classrooms. For student pre-achievement, the mean judgment accuracy score was 0.55 (SD = .17). For self-concept of ability, the mean teacher judgement accuracy score was 0.40 (SD = .21). Some teachers judged student pre-achievement quite accurately (e.g., classes 1 and 17, where accuracy was 0.71 and 0.85, respectively), whereas teachers were less accurate at judging self-concept of ability (e.g., class 10: -0.08).

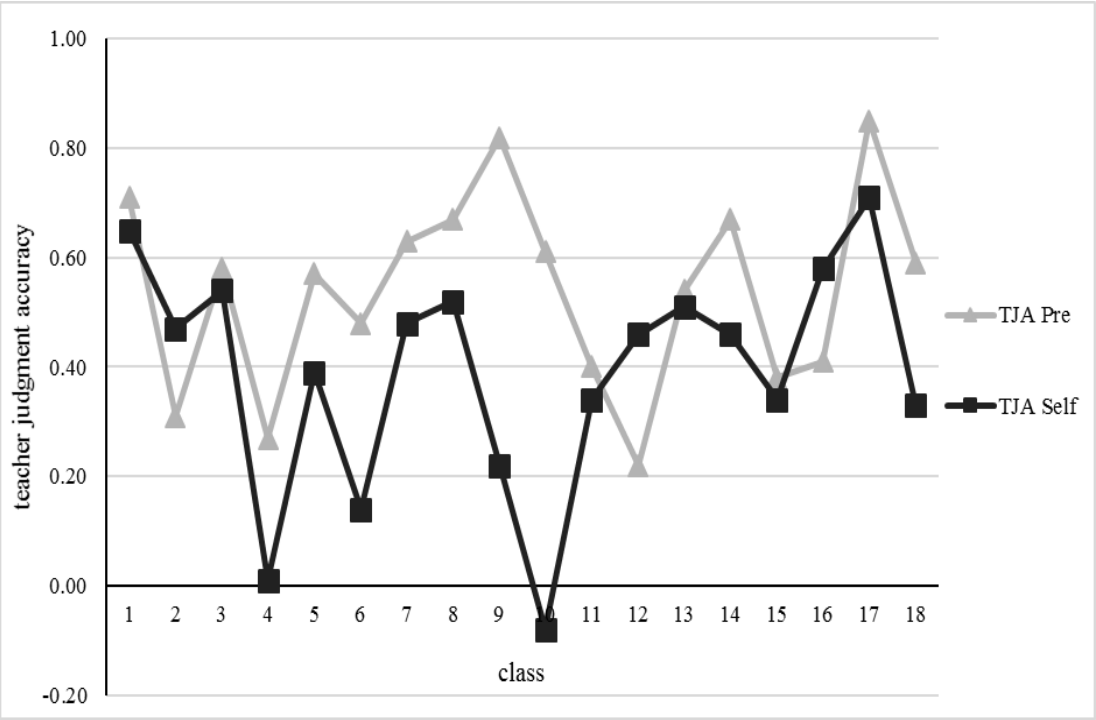

Figure 9. Teacher judgment accuracy of student characteristics in each classroom, measured as correlation between teacher and student report on class level (Pielmeier, Huber et al., 2018)

Note. TJA = Teacher Judgement Accuracy; Pre = Pre-Achievement; Self = Self-concept

7.6.1.4 Intercorrelations between Investigated Variables

In order to analyze possible relationships between investigated variables, correlation coefficients were calculated (Spearman). These analyses showed positive relationships between student pre-achievement and student self-concept of ability in mathematics ($r = .63$, $p < .01$). Furthermore, a relationship between teacher judgment accuracy of pre-achievement and teacher judgment accuracy of self-concept of ability was found ($r = .31$, $p < .01$). These findings indicate that students with high pre-

achievement are also likely to rate their self-concept of ability as high. Furthermore, teachers who are accurate in judging pre-achievement are also more likely to judge self-concept of ability accurately. Additionally, the findings showed a systematic relationship between teacher judgment of pre-achievement and student self-concept of ability ($r = .12$, $p > .05$), meaning that students with high self-concept of ability are often rated by their teachers accurately with regard to their pre-achievement.

7.6.2 Multi-level Analyses: Effects of Student Characteristics and Teacher Judgment Accuracy on Verbal Teacher-Student Interactions

As described before, this study explores to what extent teacher and student behavior in verbal interactions in classrooms can be explained both by individual student characteristics and teacher judgment accuracy as diagnostic teacher skills. Thereby, it has to be kept in mind that from the viewpoint of an individual student, instances of interactions with teachers are rather scarce events, which limits the possibilities to explain variance with regard to teacher judgment accuracy and student characteristics in this study. Nevertheless, it is of interest to explore behavioral teacher and student data in the process of teaching and learning in a classroom since important learning processes take place in this kind of educational setting.

To empirically test the second research question, two-level (level 1: student, level 2: class/teacher) analyses were applied. The results are presented for each dependent variable separately in Table 1 (verbal student engagement), Table 2 (elaborating teacher questions), and Table 3 (supportive teacher feedback). As described in the methods section, three models were run for each dependent variable. The results will be presented for each model separately.

7.6.2.1 *Student Characteristics and their Effect on Verbal Teacher-Student Interactions*

This section reports on the extent to which student characteristics can predict teacher-student verbal interactions. The analyses on student level, as shown in Column 1 of Tables 1, 2 and 3, indicate that pre-achievement did not significantly contribute to

explaining behavior in interactions, neither with regard to verbal student engagement, nor with regard to elaborating teacher questions and supportive teacher feedback.

In contrast, students' self-concept of ability showed a pattern of systematic relationships with behavior in teacher-student interactions (Column 1, Tables 1-3). Higher students' self-concept of ability was significantly and positively associated with higher verbal engagement in interactions with teachers (Table 1). Students' self-concept of ability was also positively related to interactions of elaborating teacher questions (Table 2) and supportive teacher feedback (Table 3).

7.6.2.2 *Effects of Teacher Judgment Accuracy on Verbal Teacher-Student Interactions*

This section reports on the extent to which teacher judgment accuracy as a form of diagnostic teacher skill is systematically related to verbal teacher-student interactions. Contrary to the assumptions, teacher judgment accuracies did not show any systematic relationship with the three dependent behavioral variables (Table 1-3). These findings indicate that teacher judgment accuracy did not contribute to predicting a generally higher level of verbal teacher-student interactions in classrooms.

With regard to the interplay between teacher judgment accuracy and student characteristics (statistically: interaction effects), the analyses showed the following results (Model 3, Table 1-3): Teacher judgment accuracy in interplay with the level of student pre-achievement was negatively related to the frequency of elaborating teacher questions (Table 1-3), indicating that teachers adapted interactions by providing more elaborating questions especially for students with lower pre-achievement. No such adaptive behavior could be found regarding student engagement or teacher feedback.

Regarding the effects of the statistic interplay for self-concept of ability showed the following results: Student self-concept of ability was positively related to verbal teacher-student interactions such as increased verbal student engagement, elaborating teacher questions, and (in tendency) supportive teacher feedback. Teacher judgment accuracy of student self-concept of ability on class level was not significantly related to any verbal teacher-student interactions. Teachers with higher accuracy in judging their students' self-concept of ability tended to provide more supportive feedback to those

students with high self-concept of ability (Table 3). There was no interplay (statistically: interaction effect) found between teacher judgment accuracy of self-concept of ability and students' level of self-concept of ability regarding teachers' questioning behavior and student engagement. Thus, no 'compensating' aspects of teacher judgment accuracy could be found with regard to student self-concept of ability.

Table 1. Multi-level analyses: **Student engagement** as predicted by student characteristics and teacher judgment accuracy (Pielmeier, Huber et al., 2018)

	Model 1	Model 2	Model 3
		Fixed effects	
	γ (SE)	γ (SE)	γ (SE)
Intercept	0.02 (0.10)	0.02 (0.09)	0.03 (0.10)
Level 1 Student level			
Pre	-0.08 (0.07)		-0.07 (0.07)
Self	0.19 (0.07) **		0.18 (0.07) **
Level 2 class level			
TJA Pre		0.07 (0.09)	0.04 (0.10)
TJA Self		0.07 (0.09)	0.07 (0.10)
Interactions			
Pre Stud x TJA Pre			-0.01 (0.06)
Self Stud x TJA Self			0.07 (0.05)
		Random effects	
	SD	SD	SD
Intercept	0.35	0.32	0.36
Residual	0.94	0.95	0.94
		Model fit	
Marginal R^2	0.02	0.02	0.04
Conditional R^2	0.14	0.11	0.16

Note. 0.1 +, 0.05 *, 0.01 **, 0.001 ***; Stud = Characteristic on individual student level; TJA = Teacher judgement accuracy; Pre = Pre-Achievement; Self = Self-concept

Table 2. Multi-level analyses: **Elaborating teacher questions** as predicted by student characteristics and teacher judgment accuracy (Pielmeier, Huber et al., 2018)

	Model 1	Model 2	Model 3
		Fixed effects	
	γ (SE)	γ (SE)	γ (SE)
Intercept	0.05 (0.15)	0.05 (0.15)	0.05 (0.16)
Level 1 Student level			
Pre	-0.05 (0.06)		-0.03 (0.06)
Self	0.18 (0.06) **		0.18 (0.06) **
Level 2 class level			
TJA Pre		0.06 (0.16)	0.01 (0.16)
TJA Self		-0.07 (0.16)	-0.06 (0.16)
Interactions			
Pre Stud x TJA Pre			-0.13 (0.05) *
Self Stud x TJA Self			0.07 (0.05)
		Random effects	
	SD	SD	SD
Intercept	0.58	0.60	0.63
Residual	0.85	0.87	0.85
		Model fit	
Marginal R^2	0.02	0.00	0.04
Conditional R^2	0.33	0.33	0.38

Note. 0.1 +, 0.05 *, 0.01 **, 0.001 ***; Stud = Characteristic on individual student level; TJA = Teacher judgement accuracy; Pre = Pre-Achievement; Self = Self-concept

Table 3. Multi-level analyses: **Supportive teacher feedback** as predicted by student characteristics and teacher judgment accuracy (Pielmeier, Huber et al., 2018)

	Model 1	Model 2	Model 3
		Fixed effects	
	γ (SE)	γ (SE)	γ (SE)
Intercept	0.04 (0.13)	0.04 (0.14)	0.05 (0.14)
Level 1 Student level			
Pre	0.03 (0.06)		0.04 (0.07)
Self	0.13 (0.07) *		0.13 (0.07) +
Level 2 class level			
TJA Pre		0.00 (0.14)	-0.02 (0.14)
TJA Self		0.03 (0.14)	0.03 (0.14)
Interactions			
Pre Stud x TJA Pre			-0.04 (0.06)
Self Stud x TJA Self			0.09 (0.05) +
		Random effects	
	SD	SD	SD
Intercept	0.51	0.55	0.54
Residual	0.89	0.90	0.89
		Model fit	
Marginal R^2	0.02	0.00	0.03
Conditional R^2	0.26	0.27	0.29

Note. 0.1 +, 0.05 *, 0.01 **, 0.001 ***; Stud = Characteristic on individual student level; TJA = Teacher judgement accuracy; Pre = Pre-Achievement; Self = Self-concept

7.7 Discussion

This study contributes to research on teacher judgment accuracy of student characteristics and its relevance for verbal teacher-student interactions in classrooms. Using a multi-methodological research approach, the study links research on individual student characteristics and teacher judgment skills to behavioral research on verbal teacher-student interactions in classrooms. This is the first empirical study of this kind. However, it has to be pointed out once more that from the viewpoint of an individual student, interactions with a teacher in one lesson regarding individual engagement,

teacher elaborating questions, and teacher feedback are rather scarce events. Within these scarce events, it is, however, of interest to study which factors might contribute to these interactions that have been shown to be relevant for further learning processes and outcomes (Chi, 2009; Mercer et al., 1999; Pehmer et al., 2015a; Resnick et al., 2010; Turner et al., 2003; Webb et al., 2014). In this study, two factors have been explored: the relevance of individual pre-achievement and self-concept of ability as prerequisites for learning as well as teacher judgment accuracy regarding those two characteristics. From a methodological point of view, studying scarce events leads to challenges with regard to the limited amount of variance that can be explained by factors as investigated in this study. In this sense, this study is comparable to other research addressing behavioral data (Blut, 2008; Cohen, 2010; Greene, 2012; Michaelis, 2009). Given this methodological context of the study, the findings are of exploratory character, need to be replicated by other studies and have to be interpreted tentatively.

7.7.1 Student Pre-achievement and Self-concept of Ability and their Effect on Verbal Teacher-Student Interactions

The findings of this study support previous research with regard to the fact that student characteristics influence student engagement in classroom instruction in important ways (Jurik et al., 2013, 2014; Lipowsky et al., 2007; Lipowsky et al., 2009; Pauli & Lipowsky, 2007; Pehmer et al., 2015a; Turner et al., 2011). Also found in this study, students with higher self-concept of ability are more engaged in verbal teacher-student interactions, indicated by higher verbal engagement as well as more frequent elaborating teacher questions and supportive teacher feedback (Jurik et al., 2013, 2014; Lipowsky et al., 2007; Pauli & Lipowsky, 2007; Seidel, 2006). Contrary to previous research findings, no significant effects were found for student pre-achievement when at the same time taking into account student self-concept of ability. This finding might be of importance for further research since previous studies often focused on student achievement and this indicator might have been confounded with students' self-concept of ability (Marsh et al., 2006; Marsh & Martin, 2011). In most studies investigating effects of pre-achievement on teacher-student interactions, students were grouped and compared according to high and low achievement level without controlling for additional student characteristics such as self-concept of ability (Ismail & Majeed, 2011).

It must be pointed out that observed teaching acts such as teacher questioning and feedback might be, to some extent, interrelated with the frequency of student engagement. This is because it is more likely that engaged students will directly react to elaborating questions and receive more supportive feedback. However, this interrelation actually represents the kind of complex micro learning environment that especially highly self-confident students seem to experience in their classrooms. This kind of "Matthew effect", thus, can be described once more for these self-confident students (Seidel, 2006; Seidel & Shavelson, 2007). As research shows, higher active student engagement is, in turn, positively related to students' further experience of internal learning activities, such as cognitive activation, competence, and autonomy (Furtak & Kunter, 2012; Jurik et al., 2013; Pauli & Lipowsky, 2007; Pehmer et al., 2015a; Turner et al., 2003). This might explain relevant components of the Matthew-effect (Stanovich, 1986) regarding long-term learning outcomes and school success in which often strong – or as found in this study – confident students continue to be strong and confident whereas others are being left behind (Kiemer, Gröschner, Kunter, & Seidel, 2016; Kiemer, Gröschner, Pehmer, & Seidel, 2015; Lipowsky et al., 2009).

7.7.2 Teacher Judgment Accuracy of Student Characteristics and Effects on Verbal Teacher-Student Interactions

Based on previous research and theoretical considerations (Hattie & Timperley, 2007; Kaiser et al., 2013; Ready & Wright, 2011; Südkamp et al., 2012) it was expected that teachers who judge the level of student characteristics in their classrooms more accurately would have a higher level of verbal teacher-student interactions in their classrooms. These teachers were expected to pose more elaborating teacher questions and to provide more supportive feedback to all students. In testing these conjectures, however, empirical confirmation could be found neither with teacher judgment accuracy of pre-achievement nor with teacher judgment accuracy of self-concept of ability.

It was also conjectured that teachers' diagnostic skills in accurately judging student characteristics would be particularly relevant for adaptive teaching acts (Praetorius et al., 2013). Therefore, the interplay (statistically interaction effects) between teacher judgment accuracy and student characteristic levels were additionally considered. The findings confirmed the conjectures, showing that teacher judgment accuracy and student pre-achievement level interacted significantly with regard to parts

of verbal teacher-student interactions. In particular, higher teacher judgment accuracy was related to posing more elaborating questions in classes with lower pre-achievement levels. Teachers with higher diagnostic skills may have tried to cognitively engage especially students with lower pre-achievement in order to close achievement gaps for these students (as also described by Shavelson & Stern, (1981). However, no significant interplay was observed for student engagement and supportive teacher feedback. These results tentatively support previous findings from Chin (2007) that teacher questions can be especially beneficial for students with low pre-achievement, since these questions can elicit deep thinking.

Regarding the interplay between teacher judgment accuracy and student self-concept of ability, the theoretical considerations regarding adaptive functions could not be confirmed. Higher student self-concept of ability was significantly related to higher verbal student engagement, elaborating teacher questions, and supportive teacher feedback, independent of the level of teacher judgment accuracy regarding student self-concept of ability. Thus, no compensation of occurring 'Matthew'-effects (Stanovich, 1986) could be observed in the direction that higher teacher diagnostic skills might lead to more supportive feedback especially for students with lower self-concept of ability (Hattie & Timperley, 2007).

7.7.3 Limitations of the Study

While this study presents findings on teacher judgment accuracy and student characteristics that are relevant for teacher-student interactions in the process of teaching in classrooms, several limitations have to be considered when interpreting the results. First, this study was conducted in the context of a tracked school system, and data collection was restricted to high-level mathematics classrooms. Pre-achievement was measured within this high-achieving group, but still the values were approximately normally distributed around a mean of 2.91 (see Figure 3). Also, regarding student self-concept of ability, the full range of the scale was observed in students (see Figure 4).

Second, student characteristics measured at the beginning of the school year were linked to behavior of verbal teacher-student interactions about three to four months after school began and observations were limited to one 45-minute lesson unit. Teachers focus on observing and diagnosing student prerequisites mainly at the beginning of a school

year (cf. Shavelson & Stern, 1981), which is why this measuring point was chosen for this study. However, the measurements might have been more accurate if the assessment had taken place closer to the observed teaching unit. Therefore, further research could focus on different points in time during a school year and varying situations of teacher assessments as well as adaptations of their judgments in the process of teaching a school class could be studied (Herppich et al., 2017). Furthermore, by building percentile groups in order to match teacher and student data, detailed information about the exact performance of each student got lost. Despite these limitations, first indications of possible effects of teacher judgment accuracy in interplay with student characteristics were found. Furthermore, with regard to analyzing a single lesson unit, research has shown that verbal teacher-student interactions represent quite typical and frequently occurring routines that are likely to occur the same way in other lessons throughout the school year (Praetorius, Pauli, Reusser, Rakoczy, & Klieme, 2014; Seidel & Prenzel, 2006). However, more research is also required in this field to determine with more precision how many observation units are required for valid and reliable behavioral findings representative of classroom behavior throughout a school year (Praetorius et al., 2014).

Third, another limitation might be the way variables were coded and inspected because the observed teacher behavior could be to some extent interrelated with the frequency of student engagement because it is more likely that engaged students will receive more supportive teacher feedback. As this interplay was not investigated, it might be interesting for further studies to analyze feedback in different settings, e.g. intervention studies with a professional development program where further differentiated analyses are possible.

Fourth, another limitation refers to some missing student data (N = 153), which naturally occurs when collecting data in the field and particular in ecologically valid classroom settings. Thereby, missing data was quite equally distributed across measuring points and the kind of data collected (student characteristics data, teacher judgment data, student behavior data). In this regard, no systematic bias in missing data sources or particular student group data occurred. Since estimations with regard to data sources such as behavioral student data or teacher judgments are difficult to apply in this study design, it was decided to refrain from applying additional methods such as data imputation.

Finally, possible important teaching processes such as teacher attention in the run-up to teacher acts such as questioning or feedback were not within the scope of this study. It might be that accurate teacher judgment of student characteristics does not result in specific teaching acts such as questioning and feedback but are important for regulating teacher attention processes and monitoring specific subgroups of students. Recent research applying eye-tracking methodology (Cortina, Miller, McKenzie, & Epstein, 2015; Stürmer, Seidel, Müller, Häusler, & S. Cortina, 2017; van den Bogert, van Bruggen, Kostons, & Jochems, 2014; Yamamoto & Imai-Matsumura, 2015) provides valuable insight into those teacher professional skills. This method may be particularly useful for further integrating classroom research with research on teacher diagnostic skills.

7.7.4 Conclusion

The results of this study support that student self-concept of ability is particularly relevant when explaining student engagement in verbal teacher-student interactions in classrooms. This effect was not compensated by higher levels of teacher judgment accuracy. However, higher teacher judgment accuracy of pre-achievement was related to more elaborating teacher questions for students with lower pre-achievement levels. These results provide first tentative insights into possible adaptive functions of teacher judgment accuracy that might be of relevance for further studies on teacher diagnostic skills in the process of teaching and learning in classrooms. In this sense, adaptive functions might expand the field of teacher diagnostic.

8 Study II: The Changeability of Classroom Discourse (through TPD) and its Effects on Teachers Judgments and Student Self-concept of Ability

8.1 Addressed Research Questions

In the present study, it will be empirically explored by means of case analyses how three teachers participating in a video-based TPD program on productive classroom discourse undertook changes in classroom discourse practices and developed their students' self-concept of ability and their own teachers' judgments of this student characteristic. In order to describe these change patterns, teachers were selected who had started the investigated TPD program and who had low scores for productive classroom discourse. It was assumed that in such a context, changes in both the teaching practices and classroom experiences of both the teachers and students might be observed more directly. In the present study, the following research questions were addressed:

1) How do 'low-performing' teachers change their classroom discourse practices through a TPD program?

2) Which consequences from the change in classroom discourse can be observed for
   a) student experiences and reports of their self-concept of ability and
   b) the individual teacher judgment accuracy of students' self-concept of ability?

8.2 Study Context: TPD Program the Dialogic Video Cycle

The investigated teachers participated in the video-based TPD program entitled the "Dialogic Video Cycle" (DVC). AS previously explained (see chapter 3.3), the DVC follows the heuristic of lesson planning–lesson teaching–lesson reflection inspired by the problem solving cycle (PSC) (Koellner et al., 2007). At the beginning of the intervention, the teachers met for an organizational session where they were able to meet the facilitator and research team as well as the other participants. Teachers learned about the research

involved in the TPD program and set an appointment for the pre-videotaping in their classroom, for which they had not yet had a theoretical input. After all of the teachers' pre-video recordings, the group met for a first lesson-planning workshop where they were asked to bring a concrete lesson plan. At the beginning of the session, they received theoretical input from the facilitator on productive classroom discourse by means of *talk formats* (varying student conversation settings) and *talk moves* (teacher conversational strategies), e.g. on elements of dialogic teaching such as making the learning goal clear. After the first workshop, the teachers agreed on an appointment for videotaping their planned and modified lesson. For the second and third workshop, the facilitator systematically selected representative video clips (two or three minutes each) from each teacher for collective reflection. Before watching the clips, the facilitator established a policy for the discussions of the classroom videos. Based on the concept of teacher professional vision (Stürmer & Seidel, 2015), the teachers were asked to first describe their observations (without judging), explain their colleague's decision making, and predict the students' learning. Being familiar with the video-observation policy, the group would watch a clip (of one of the group members) and the teacher on the screen would provide further explanations or contextual information (for the facilitation moves, see Alles et al., 2018). During the next step, guiding questions regarding productive classroom discourse (e.g., "How does the teacher activate his/her students?") served as the basis for providing feedback (including solutions and alternatives) or asking more questions. The reflection workshops provided a good opportunity for teachers to reflect not only on their own actions but also on the students' reactions and behaviors. Together, the teachers reflected on why some things worked and others did not and on how some students could have benefitted better. Being involved in both a reflection on both one's own teaching and the teaching of others further sensitizes a teacher on how students react. This is especially the case in regard to the talk moves aim of gaining a better understanding of student characteristics as a prerequisite and supporting students individually. The same procedure for the DVC was repeated during the second half of the academic year. Lastly, all the teachers were videotaped at the end of the school year for a post-measurement point.

To ensure the quality of the program, a feasibility check was undertaken by two independent raters who watched all of the videotaped TPD workshops. All of the workshops were screened for the fulfillment of Desimone's (2009) quality features of TPD (Gröschner et al., 2015), leading the presented study to quality-ensured results.

8.3 Sample: The Selection of Three Cases

This study involved an in-depth analyses of classroom videos and questionnaire data from a sample of the selected teachers based on the criteria that follows. First, since the development of the teachers' accurate knowledge of their students depends on the time they have been teaching a class prior to the TPD participation, the amount of time that the teachers have spent with their students should be comparable. Therefore, it was determined that the selected teachers and their classes should have known each since the start of the school year when the TPD program took place, meaning that the teachers and students did have some history together but that this history had started at a comparable point of time (about one month prior to TPD participation). This criterion accounted for the largest group of participating teachers. Second, since the possibilities for interactions between teachers and students also depend on class size, the cases should be selected from among those with a typical and comparable class size. Third, the teachers were selected from among those with a comparably low performance at the beginning of the school year in regard to classroom discourse. 'Low performers' were chosen because they have the most room for development, and might also provide more salient opportunities for these teachers to experience changes in classroom discourse practices and learn more about their students and their characteristics as prerequisites for education.

Applying these criteria led to the choice of "Lisa", "Steve", and "Daniel", three participating science and mathematics teachers (see Table 4). Daniel, the mathematics teacher, was older than the other two and had slightly more teaching experience.

Table 4. Sample Study II - Case Study (Pielmeier et al., submitted)

syno-nym	**gen-der**	**subject**	**teaching expe-rience**	**class size**	**teacher age**	**student age**	**female**	**grade**
Lisa	female	Science	3	26	32	14.58	38.5%	9
Steve	male	Science	2	24	31	15.38	25.0%	10
Daniel	male	Math	7	21	45	14.43	33.3%	9

8.4 Research Design

The data for this study was collected during the 2016-2017 academic year. The study was based on a longitudinal multi-method research design to allow for an in-depth investigation of the development of both teachers and students. At the beginning and at the end of the school year, the students' self-concept of ability was reported through a student questionnaire (see Figure 10). Teacher judgments of this student characteristic were also collected at the same measuring point (pre: at the beginning; post: at the end of the school year). Furthermore, at the pre- and post-test, one lesson was video-taped per classroom in order to assess productive classroom discourse.

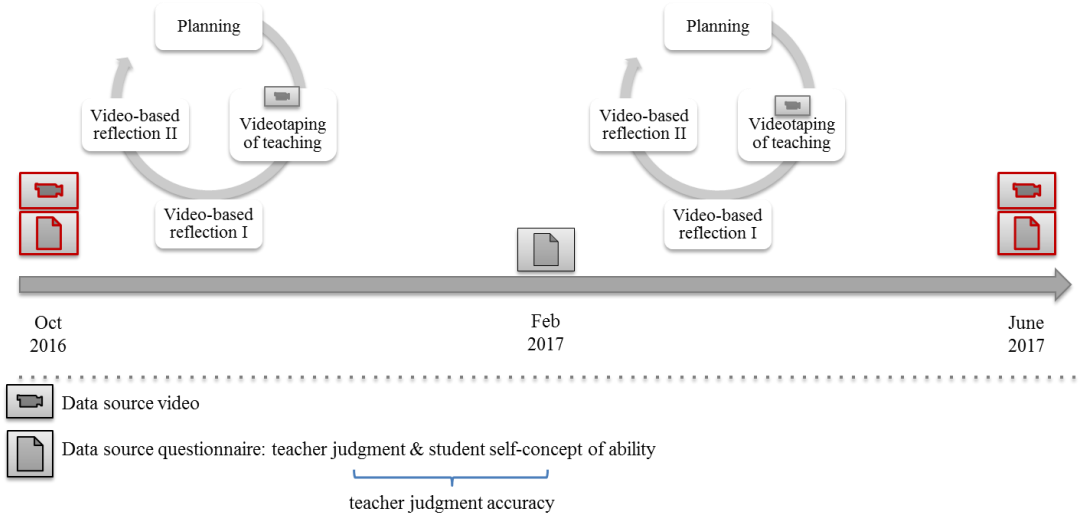

Figure 10. Study Design and Data Sources (Pielmeier et al., submitted)

8.5 Instruments

8.5.1 Video Data: Assessing Productive Classroom Discourse

Video Analysis Units. The videotaped lessons were about 40–45 minutes long (as this is the typical class length for the study country). The camera perspectives included both, an overview camera showing all students and a teacher camera that constantly followed the teacher's movements and recorded the teacher's speaking with a small fine microphone which was fixed at the teacher's neck/chest. All participants

(teachers, students) agreed to participate on a voluntary basis. Analysis units were the first 4 x 10 min segments, starting at the teacher's opening of the lesson (e.g., "Good morning"). Time sampling was chosen due to an economic perspective and to orient the study with the current analysis approaches in the field, with analysis units reaching from 10 minutes (Seidel et al., 2005) up to 20 minutes (Reznitskaya et al., 2016). Interact software was used (Mangold, 2014).

Video Rating Procedure. The quality of the classroom discourse was assessed using a theory-based coding manual with four quality dimensions for productive classroom discourse (QD 1 structured and purposeful, QD 2 activating and open, QD 3 interactive and cumulative, QD 4 supportive and scaffolding) (Schindler, Böheim, Pielmeier, Gröschner, & Seidel, 2018). Each quality dimension was specified by two or three items (Table 5). The video manual was developed based on previous research findings on classroom discourse as described in the theoretical background (see Chapter 2.3). For each segment, the enactment of the talk moves belonging to quality dimension 1-4 was rated on a six-point Likert scale.

In order to illustrate the results for all quality dimensions in a comparable and comprehensive way, all scores were translated in the format of % of the maximal achievable score (performance). For example, a teacher who had a mean teacher score of 2.4 (on the 0-5 scale) received an achievement score of 0.48, meaning 48% of the maximal performance across all rated items.

Reliability. To assure a reliable application of the described ratings, a well-documented and structured analysis of both the training and main study videos was undertaken. A master rater (> 5 years of experience in analyzing classroom discourse) rated all videotapes and was blind to the knowledge of the measurement points of the videotaped lessons (for more details, see Authors et al., 2018). All video segments were double coded by the master rater and a second rater. The master rater's results were compared to the second rater of the video material. Reliabilities ranged between .64 and .83 and could be judged as good to excellent (Wirtz & Caspar, 2002).

Table 5. Overview on video rating items to capture the classroom discourse quality (Pielmeier et al., submitted)

Quality dimension	**Item**	**Teacher item** ***"If needed…"***
QD 1 Structured and purposeful	1.1 GOAL	*"…the teacher provides a learning goal to allow for purposeful talk."*
	1.2 STRUCTURE	*"…the teacher provides a clear structure for talk so students can follow."*
	1.3 CLARITY	*"…the teacher makes expectancies and rules for talk such as active listening and verbal contributing clear."*
QD 2 Activating and open	2.1 VERBAL ACTIVE	*"…the teacher encourages students for verbally contributing to talk."*
	2.2 VERBAL PASSIVE	*"…the teacher demands to actively listen to others' contributions."*
	2.3 INITIATION	*"…the teacher provides room for elaborations, student ideas, questions and thoughts by initiating talk by e.g. open questions."*
QD 3 Interactive and cumulative	3.1 FOLLOW	*"…the teacher follows-up on students' answers and presses for more explanations, ideas, questions and thoughts."*
	3.2 LINK	*"…the teacher links answers in the classroom by connecting students' contributions or encouraging complements and critique."*
QD 4 Supportive and scaffolding	4.1 SCAFFOLD	*"…the teacher provides scaffolding for students by e.g. feedback on their learning processes or hints for further learning."*
	4.2 MISTAKE	*"…the teacher provides room for mistakes, uses them as learning source and topic of discussion."*

8.5.2 Student Subject-specific Self-concept of Ability

A student questionnaire from Marsh et al. (2005) was used to measure student self-concept of ability pre- and post-test. The scale included five items (α_{pre} = .91; α_{post} = .91) on a four-point Likert Scale. An example is: "Math/science just isn't my thing." (Range: 0 = strongly disagree to 3 = strongly agree). For each subject, the items were modified by replacing each subject accordingly.

8.5.3 Teacher Judgment Accuracy

Teachers rated each individual student in their class on the student's level of subject-specific self-concept of ability using a five-point Likert scale (0 = low, 1 = rather low, 2 = intermediate, 3 = rather high, 4 = high) (Pielmeier, Huber et al., 2018). The following instruction was given to the teachers: "Please assess the subject-specific self-concept of ability of your students on a scale from 0 to 4 (0 = low, 1 = rather low, 2 = intermediate, 3 = rather high, 4 = high). To do this, enter your rating from 0 to 4 in the marked field for each individual student." These ratings were based on the teachers' own perceptions of the students in their classrooms. To match the distribution of the students' self-report scores of self-concept of ability with the teachers' judgments of the students' self-concept of ability, the teacher judgments were grouped into four percentile groups (bottom percentile = 0 = low self-concept of ability, second percentile = 1 = rather low self-concept of ability, third percentile = 2 = rather high self-concept of ability, top percentile = 3 = high self-concept of ability).

For each student, the difference between the teacher's judgment of the student's self-concept of ability and the student's reported self-concept of ability was calculated (Thiede, Brendefur, Carney, Champion, Turner, Stewart, & Osguthorpe, 2018). This led to positive and negative values, with negative values indicating that the teacher underestimated the student and positive values indicating an overestimation. These absolute values were further used to calculate a mean for each class and subgroup within each class (high, medium, low). Values near "0" expressed a close match between the teacher and students and, therefore, high teacher judgment accuracy, whereas larger values expressed low teacher judgment accuracy (Thiede et. al, 2018).

8.5.4 Data Analysis and Illustration

Descriptive analyses were used to visualize the development that occurred during the TPD program over the school year for both for the video and questionnaire data. In each case, the video data is illustrated in one figure showing the percentages of the maximal performance for the pre- and post-measuring points for each quality dimension (Figures 11a, 12a, 13a).

Within each class, the students were split into three groups based on their self-reported level of self-concept of ability at the pre-test. This was used as an entry level in order to track the effects of the development of classroom discourse quality on the different subgroups. Therefore, three percentile groups were created. Afterwards each teacher judgment accuracy was calculated separately for the three percentile groups in each class based on the entry level of students' self-concept of ability. An individual accuracy level was then calculated as mean for each subgroup and each measuring point. For comparing the classes, the absolute values of the individual judgment accuracy scores were reported.

Taking into account subgroups based on the entry level of the students' self-concept of ability is reasonable because students with different levels of self-concept of ability behave differently (Wigfield & Wagner, 2005). There is one figure for each class (Figures 11b, 12b, 13b) showing student self-concept of ability for the pre- and post-measuring point. In additional figures (Figures 11c, 12c, 13c), the individual teacher judgment is illustrated.

8.6 Results

8.6.1 Case 1: Lisa

Lisa's classroom discourse at the beginning of the school year had a typical closed IRF pattern. By the end of the TPD program, her classroom discourse was far more open and included many of the elements of productive classroom discourse. The lines for each quality dimension in Figure 11 show the increases in Lisa's classroom discourse performance (Figure 11a). Through elements of dialogic teaching, such as goal-orientation and structure (QD 1), she encouraged her students to elicit their ideas, thoughts, and opinions (QD 2). Furthermore, student contributions were understood as resources and linked together (QD 3). A warm and welcoming learning atmosphere was created, and students received appreciation for their contributions and were given room for mistakes (QD 4).

When looking more closely at the development of Lisa's students, the student questionnaire data indicated that Lisa's students with a low entry level of self-concept of ability particularly profited from the changed classroom discourse (Figure 11b). In addition, her students with a moderate entry level of self-concept of ability showed a slight increase in their self-concept of ability. Students with a high entry level of self-concept of ability seem to experience a slight decrease.

Lisa's implementation of the productive classroom discourse elements appears to have provided her with new opportunities to better get to know and learn about her students. This is shown in Figure 11c. Lisa's teacher judgment accuracy of individual students' self-concept of ability improved, especially for students with a moderate entry level of self-concept of ability. Interestingly, Lisa did not improve her individual teacher judgment accuracy for students with a high entry level of self-concept of ability; there was a marginal decrease. At the beginning of the school year Lisa, was most accurate in judging students with a low entry level of self-concept of ability, but there was a slight decrease in accuracy at the end of the school year.

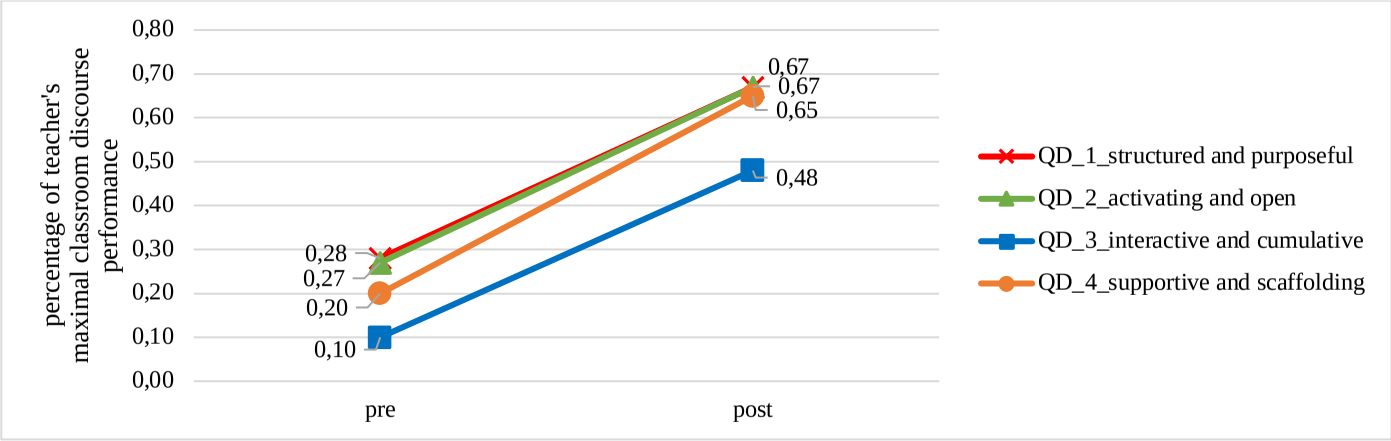

Figure 11a. Video data in Lisa's class showing percentages of teachers' classroom discourse performance for each quality dimension (QD) (1-4) for pre and post measuring point.

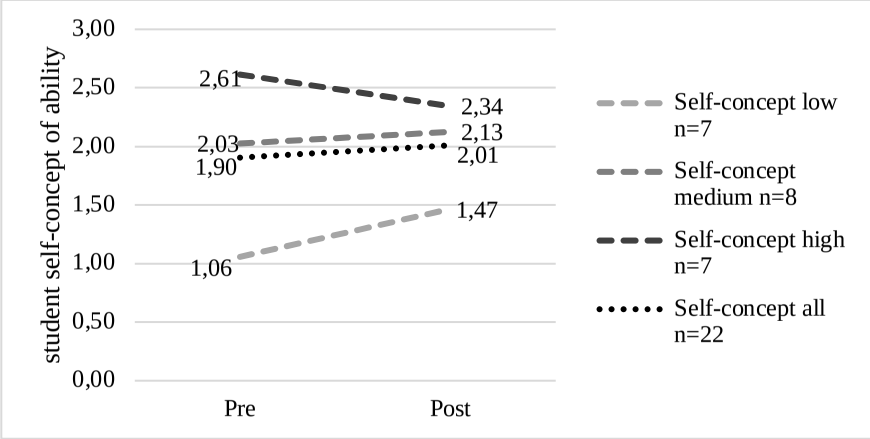

Figure 11b. Student self-concept of ability in Lisa's class

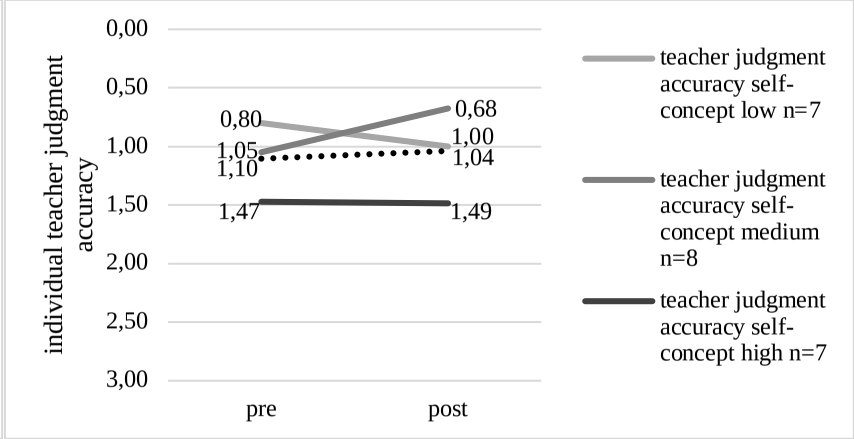

Figure 11c. Development of Lisa's individual teacher judgment accuracy

Figure 11. Data Case Lisa (Pielmeier et al., submitted)

Note. Only students who attended all measurement points are included.

8.6.2 Case 2: Steve

Steve also started with a low level of classroom discourse performance in all four quality dimensions (Figure 12). Over the school year (Figure 12a), Steve structured his classroom discourse more and made learning goals explicitly clear, which is relevant for students as they can integrate the learning goal for themselves and follow the classroom discourse better. In addition, the more noticeable structure of the discourse seemed to have led to more frequent meaningful student contributions (QD 1). Students had more opportunities to contribute to the classroom discourse as active participants. In addition, Steve paid closer attention when students were actively listening (QD 2). All this was possible due to opening the classroom discourse by, for example, asking open-ended questions and giving students time to elaborate on answers. This quality aspect of productive classroom discourse was also characterized by appreciating student responses and linking student responses to each other (QD 3). Further, the changed classroom discourse had improving ratings for a supportive learning environment with room for mistakes and supportive feedback (QD 4). Overall, Steve improved his classroom discourse performance.

Although there was a positive change in the classroom discourse, there were almost no changes in his students' self-concept of ability (Figure 12b). Students with low entry level of self-concept of ability stayed over the entire school year on a constant level. Students with a moderate entry level of self-concept of ability showed a slight decrease. Students with high entry level of self-concept of ability (which was only slightly differentiated from those with a medium level) decreased to the same level as the medium level group. It seems like Steve's students did not perceive or profit from the changes in the classroom discourse with regard to their self-related judgments of their capabilities.

Although Steve's performance did not increase as much as Lisa's, it still seems as if Steve used the newly opened opportunities in his classroom discourse to improve his judgments about his students' self-concept of ability. In his case, there were only a few differences in his teacher judgment accuracy for the three subgroups of students (Figure 12c). He was fairly aware of students with low entry level of self-concept of ability and showed a marginal improvement by the end of the school year. Regarding the individual teacher judgment accuracy for students with moderate entry level of self-concept of ability, there was a slight improvement by the end of the school year. There was also a slight improvement of the individual teacher judgment accuracy found for students with a high entry level of self-concept of ability.

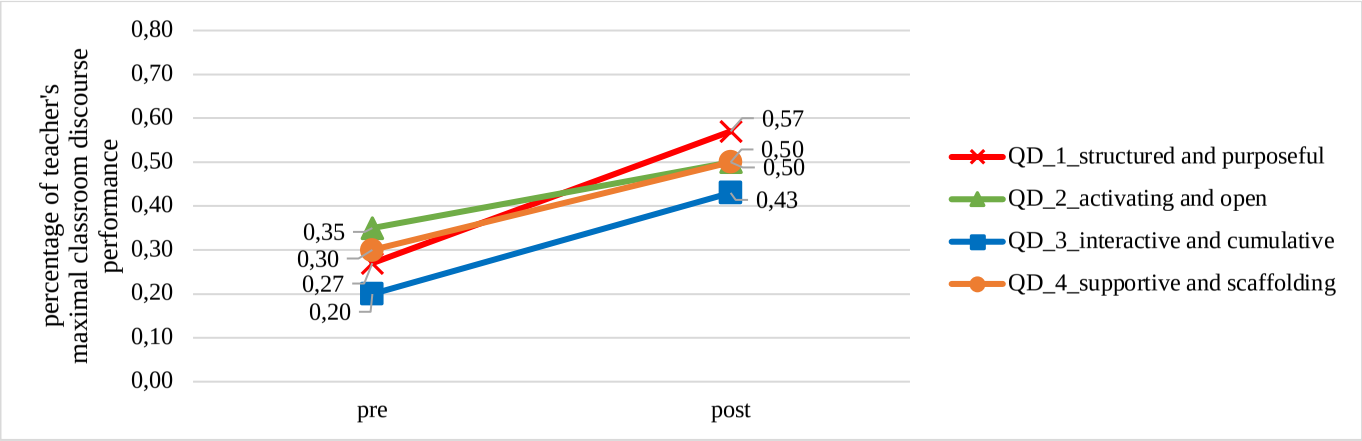

Figure 12a. Video data in Steve's class showing percentages of teachers' classroom discourse performance for each quality dimension (QD) (1-4) for pre and post measuring point.

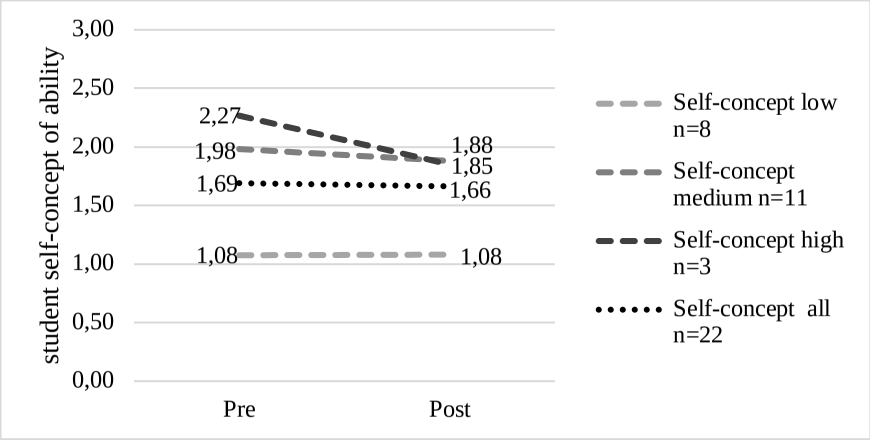

Figure 12b. Student self-concept of ability in Steve's class

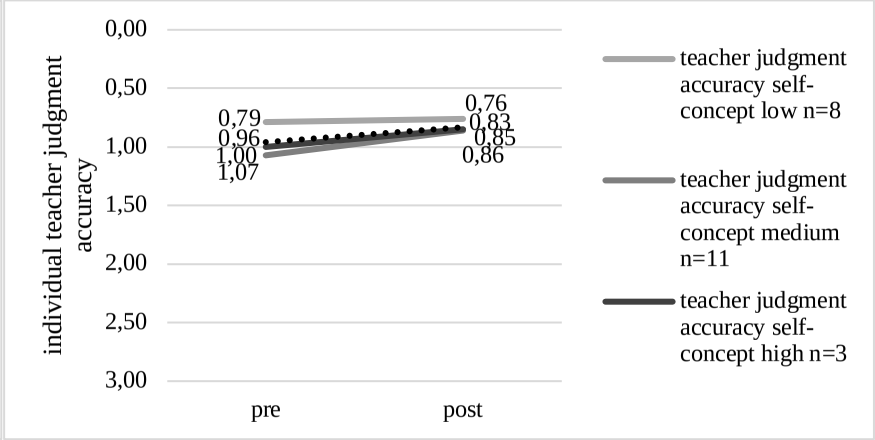

Figure 12c. Development of Steve's individual teacher judgment accuracy

Figure 12. Data Case Steve (Pielmeier et al., submitted)

Note. Only students who attended all measurement points are included.

8.6.3 Case 3: Daniel

Daniel is a teacher who also started with a low classroom discourse performance for all four quality dimensions (Figure 13a). By the end of the school year, Daniel performed at a much higher level.

Daniel's students all profited from the changed classroom discourse. Students with a low and moderate entry level of self-concept of ability showed an increase in their self-concept of ability over the school year (see Figure 13b). Students with a high entry level of self-concept of ability showed a slight increase. In Daniel's class, the student subgroups were obvious and remained separate, but within each subgroup, there were changes indicating that through the changed, open, and more productive classroom discourse, all three subgroups could change within their ability.

It appears that the newly opened 'windows' in which Daniel had the chance to get to know his students better during the classroom discourse were especially valuable for students with low and medium entry levels of self-concept of ability. Daniel's individual teacher judgment accuracy (Figure 13c) improved, especially for students with low and moderate entry levels of self-concept of ability. At the beginning of the school year, Daniel was more aware of students with a high entry level of self-concept of ability, but by the end of the school year he could not judge these students accurately.

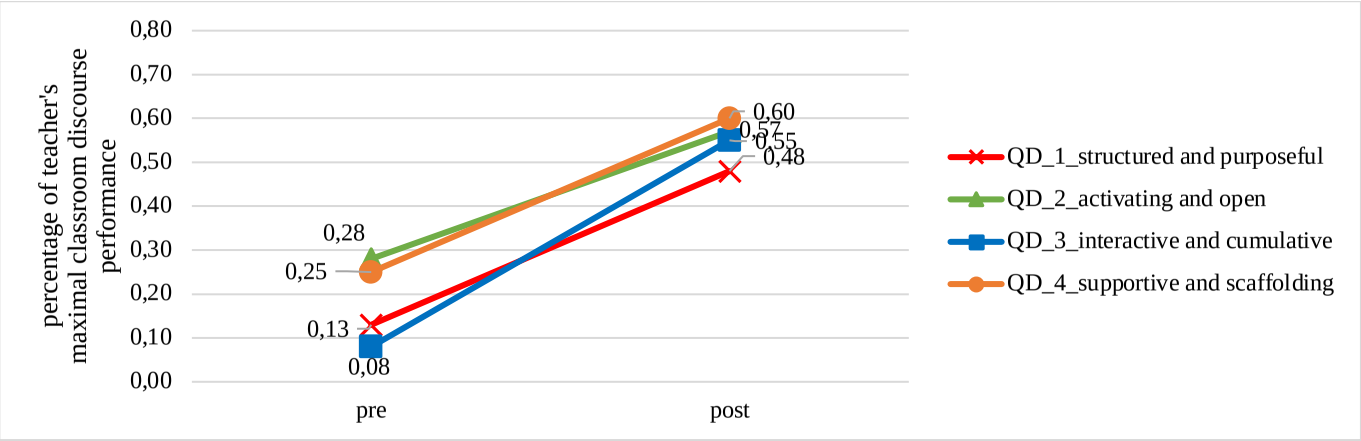

Figure 13a. Video data in Daniel's class showing percentages of teachers' classroom discourse performance for each quality dimension (QD) (1-4) for pre and post measuring point.

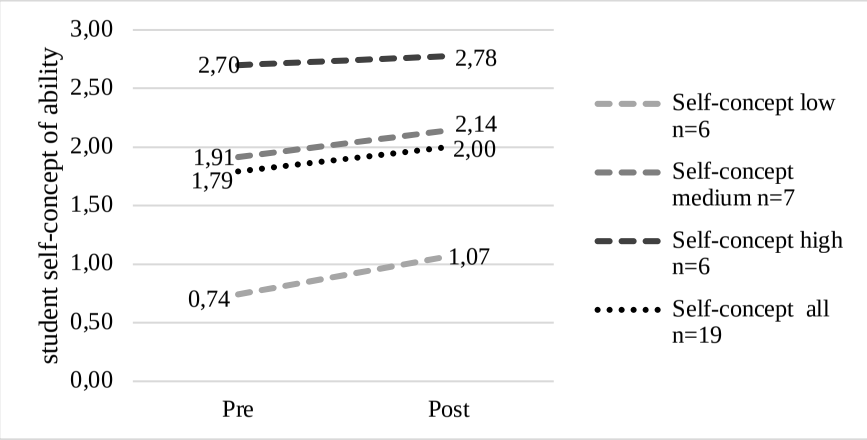

Figure 13b. Student self-concept of ability in Daniel's class

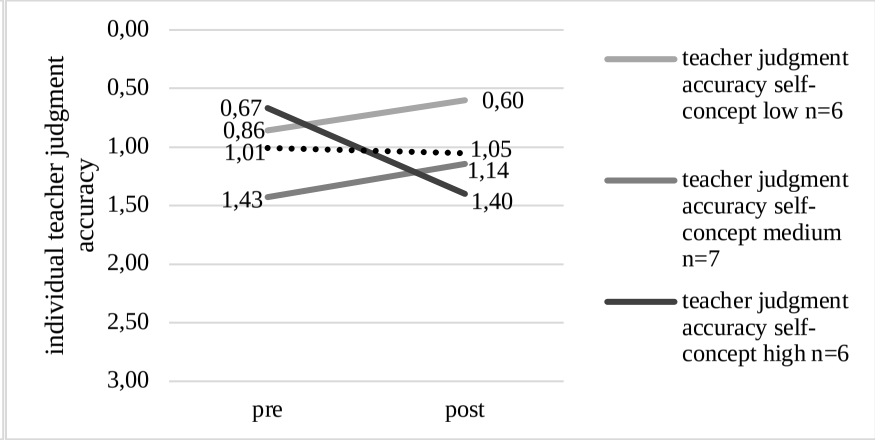

Figure 13c. Development of Daniel's individual teacher judgment accuracy

Figure 13. Data Case Daniel (Pielmeier et al., submitted)

Note. Only students who attended all measurement points are included.

8.7 Discussion

This study has presented three case analyses of math and science teachers who participated in the Dialogic Video Cycle (DVC), a video-based TPD program that aims to improve productive classroom discourse. The goal of this study was to illustrate how changes in the classroom discourse of individual teachers is linked to different experiences for learning about student prerequisites. Therefore, this study investigated how teachers who were low-performing at the beginning of participation changed their classroom discourse practices throughout a TPD program (research question 1) and whether these changes coincided with changes in the students' self-related judgments of their own capabilities (research question 2a). In addition, this study also explored whether teachers improved in accurately judging their students' self-concept of ability (research question 2b). Thus far, only a few relationships between teaching practices and the knowledge of students' individual characteristics have been examined (Alonzo & Kim, 2018; Machts et al., 2016; Sedova & Salamounova, 2016). Current research on classroom discourse has not yet investigated teachers' individual changes in classroom discourse quality in connection to a changed judgment of student characteristics such as a student's self-concept of ability. Teachers with a high potential for change were investigated in-depth in order to get an idea of how they implemented the contents of the TPD and how these contents might have resulted in learning more about their students and their characteristics as prerequisites for successful learning.

8.7.1 Classroom Discourse Performance

All three teachers started with a comparably low level of productive classroom discourse. Throughout participation in the TPD program, they were able to make a positive change to their classroom discourse practices across all quality dimensions. Productive classroom discourse is characterized by the previously introduced quality dimensions. One element is that classroom discourse has to be goal-orientated and structured (quality dimension 1) (Resnick et al., 2010; Seidel & Prenzel, 2006) so that students can contribute in a meaningful way. Furthermore, through open-ended teacher questions students can elicit their ideas, thoughts and opinions (quality dimension 2) (Chi, 2009; Oliveira, 2010). It is also necessary for the different student responses to be linked and understood as a resource for the ongoing discourse (quality dimension 3)

(Gomez Zaccarelli et al., 2018; Osborne et al., 2016). In addition, a warm and welcoming learning atmosphere should be created, where student answers are also treated as a resource for the ongoing discourse (Gomez Zaccarelli et al., 2018) and there is room for and appreciation of mistakes, rather than an atmosphere of pure evaluation (Grassinger et al., 2018). This should include feedback containing information on the students' further learning processes (quality dimension 4) (Hattie & Timperley, 2007). The differential picture showed that Lisa demonstrated the strongest changes among the three analyzed cases. Steve showed the strongest changes regarding quality dimension 1 (structured and purposeful). Daniel showed the strongest changes for his initially weakest quality, dimension 3 (interactive and cumulative). All these findings are in line with previous studies that have shown that teachers are able to successfully implement talk moves and aspects from classroom discourse TPD programs but that they do not implement all talk moves to a similar extent (Gomez Zaccarelli et al., 2018; Michaels et al., 2008; Pehmer et al., 2015a, 2015b; Schindler, Gröschner, & Seidel, 2016; van de Pol et al., 2017; Vrikki et al., 2017). It would now be interesting in to see if in the longer term, the other quality dimensions of Steve's and Daniel's classroom discourse would follow with improvements, perhaps if they joined another TPD or were involved in a follow-up study. It would also be interesting to see if Lisa can improve further and hold the high level of her changed classroom discourse for all quality dimensions and in other classes.

8.7.2 Student Self-concept of Ability

Student self-concept of ability is known as one of the key characteristics for student participation, which is highly relevant for high-quality classroom discourse (Dai & Sternberg, 2004; Marsh & Martin, 2011; Seidel & Shavelson, 2007). The students' self-concept of ability showed positive changes in all three cases. In the three subgroups (the low, medium, or high entry level for self-concept of ability) of two classes (Lisa's and Daniel's class), students with low and moderate entry levels for self-concept of ability showed a slightly higher self-concept by the end of the school year. With exception of Daniel's class, the subgroup of students with a high entry level for self-concept of ability decreased slightly by the end of the year. This may be to some extent due to ceiling effects. Previous research has shown that students with a high self-concept of ability are used to being more engaged (Blöte, 1995; Jurik et al., 2013, 2014; Pauli

& Lipowsky, 2007), resulting in them receiving more supportive feedback (Blumenfeld et al., 1982; Ismail & Majeed, 2011). Both aspects might have become less due to the changed teacher-student interaction, which might have especially benefitted students with a lower entry level of self-concept of ability, but at the same time, caused a drop for students with a high self-concept of ability. Overall it seems like through the changed classroom discourse in all three cases, there were more opportunities for students to contribute, and as a consequence, their self-concept of ability increased.

8.7.3 Teacher Judgment Accuracy

Students contributions are not only highly relevant for the success of a productive classroom discourse (Chi, 2009; Mercer et al., 1999; Resnick et al., 2010; Webb, 2009) but also for teachers to get to know their students better. Due to the changed classroom discourse, it seems like there were more opportunities for the teachers in the TPD program to better judge their students' characteristics. Through a more intense interaction, teachers had a greater opportunity to learn about their students' individual characteristics, thinking, and learning processes from verbal contributions, and this is the main source for teachers. Teachers have difficulties in judging student self-concept of ability accurately (Praetorius et al., 2013; Praetorius & Südkamp, 2017; Spinath, 2005; Südkamp & Praetorius, 2017). In particular, this is true for students with a low self-concept of ability as these students are less involved in the discourse (Jurik et al., 2013; Pehmer et al., 2015b). Consequently, fewer verbal contributions mean that there are fewer opportunities for teachers to understand student thinking and learning processes and to make accurate judgments of their students. In the three analyzed cases, it seems like a positive change in classroom discourse affected the teachers' perception of their students, in some cases resulting in better teacher judgment accuracies of students' self-concept of ability by the end of the school year. The findings are in line with previous research showing that changes in teaching practices through TPD also include the potential for teachers to get to know their students better. For example, Alonzo and Kim (2018) showed that through elaborating questions, higher classroom discourse quality was supported, and the higher discourse quality was related to the quality of teachers' judgments.

The teachers might have become better in judging their students with a low self-concept of ability probably because of the more intensive teacher-student interaction

during the classroom discourse. In order to adapt teaching routines to student needs, it is a professional demand that teachers can accurately judge their students' characteristics (Corno, 2008). Student participation is highly relevant for teachers in obtaining opportunities to get to know their students better through verbal contributions. Brühwiler (2017) showed that student participation was significantly related to teacher judgment accuracy. This finding is supported by the present study, in which it was showed that, in all three cases, students participated more in the classroom discourse by the end of the school year (see quality dimension 2) and teacher judgment accuracies also improved for at least some subgroups over the school year.

The results are also in line with the previous findings from Alonzo and Kim (2018), who showed that judgment quality and discussion quality were closely related and that teacher expectations and judgments are to an extent related to the teacher's discourse behavior (Machts, Kaiser, Schmidt, & Möller, 2016; Sedova & Salamounova, 2016; Sedova, Sedlacek, & Svaricek, 2016). Furthermore, TPD programs and research on classroom discourse should also address teacher judgments and how to improve them. Some light has been shed on this relevant aspect for teacher learning and on the role of teachers' judgment during verbal interactions with students and students' learning about student characteristics.

8.7.4 Limitations and Directions for Further Research

While this study has presented findings on individual developments regarding teachers` classroom discourse performance, students' self-concept of ability, and individual teacher judgment accuracies, several limitations have to be considered when interpreting the results and reading the cases. This study was conducted in a volunteer TPD program at German high-track schools. Furthermore, only one lesson was videotaped at the beginning and one at the end of the school year, and therefore, additional video analysis throughout the school year might yield additional information on the individual trajectories of practice changes. In addition, mixed method approaches also including interview data might yield further insight into both the teachers' and students' perceived changes in their classrooms. By combining these data sources, a further understanding of verbal teacher-student interactions, student and teacher perceptions, and their link to teacher judgment can be obtained.

8.7.5 Conclusion

In this case study, the individual changes of three teachers who participated in a year-long video-based TPD program were illustrated in-depth. These findings provide insights into the relationships between the individual changes teachers make in their classroom discourse practices and the student perceptions of their abilities in these classrooms along with the ability of teachers to judge this perception. The findings tentatively indicate that the implementation of productive classroom discourse elements resulted in positive changes in student behavior, particularly for students with a low self-concept of ability. The described cases illustrate important ‘side’ effects in TPD research that indicate how the ecology of a classroom environment in which teachers and students interact can be affected by external input, such as TPD. These effects should be considered more systematically in future TPD research.

9 Discussion

9.1 Overview and General Discussion of Central Findings

The aim of this dissertation was to get a better understanding of the interplay of classroom discourse and teacher judgment accuracy in relation to relevant student characteristics. This chapter provides an overview of the central findings of Study I and II. These findings are discussed together and with regard to the previously introduced theoretical background in Chapter 2 to 5. Therefore, the framework model of this dissertation (see Figure 1) and the three central research questions which guided this dissertation (see Chapter 6.1) should be kept in mind: First, the results regarding classroom discourse will be discussed. Second, student characteristics in the context of classroom discourse will be discussed. And third, teacher judgment accuracies of student characteristics will be discussed in the context of classroom discourse. The discussion is always with regard to Study I and II in the light of the conjectures and current research.

9.1.1 Classroom Discourse Practices

First, it is discussed how classroom discourse looked like in Study I and how it changed in Study II as these findings are the basis for the discussion of the other findings of this dissertation. Chapter 7.7 discussed the effects of student characteristics and teacher judgment accuracies on certain classroom discourse patterns. So far, the status quo of classroom discourse patterns has not yet been discussed on its own. As it serves as basis for Study II, which focused on the changes of classroom discourse through a video-based TPD program, these findings are now discussed within the context of this dissertation and in relation to current research. Overall, the findings of this dissertation are in line with previous findings from research in this field.

In Study I, there were many (about 73) content-related student contributions, indicating that these were probably more likely short answers. Previous studies have shown that especially in physics instruction students mainly provide keywords (Seidel & Prenzel, 2006). Student verbal engagement is largely influenced by teacher's

questioning behavior (Cazden, 2001). In Study I only elaborating questions have been investigated, meaning questions which invite students to think deeply (Cazden, 2001; Pehmer et al., 2015a) and encourage student to reproduce known facts, link new information to previous existing knowledge (e.g. Chin, 2006). In other words, questions which longer and elaborated student answers would be expected as these were not low cognitive level questions like reproducing questions. This also explains on the one hand the comparable few questions (about 23) in relation to student engagement. Following the typical IRF pattern, teacher feedback was the third investigated discourse pattern. Regarding teacher feedback, the focus was on supportive teacher feedback, meaning feedback containing constructive and supportive information (Hattie & Timperley, 2007) as this has been shown to have positive effects on students learning processes (Gan & Hattie, 2014; Kobarg & Seidel, 2007; Timperley, 2013) and also positive effects on the development of motivational-affective student characteristics, e.g. self-concept of ability (Deci et al., 1999; Jurik et al., 2014). Again this was a scarcer event as not the evaluative feedback, which is most common in the typical classroom discourse (Chafi & Elkhouzai, 2016; Hattie & Timperley, 2007), was investigated. Results of Study I indicate that only few student contributions received supportive teacher feedback and also elaborating questions were not the main questioning type, which should be ideally.

These results of Study I support the call for teacher professional development in the context of classroom discourse (see also (Hennessy et al., 2016)). Study II adds on to this. A video-based teacher professional development program, the Dialogic Video Cycle, was implemented and investigated. In a first step, the changes regarding classroom discourse will be outlined and discussed with regard to other TPD programs and their effects.

To get an understanding of how teachers implement instructional practices a deep look is required (Gomez Zaccarelli et al., 2018). Therefore, a case study was conducted in Study II. All three teachers showed positive changes of their classroom discourse performance post-TPD participation. All three teachers started the TPD program with a comparably low classroom discourse performance. Regarding all four quality dimensions (QD 1 structured and purposeful, QD 2 activating and open, QD 3 interactive and cumulative, QD 4 supportive and scaffolding) the three analyzed teachers showed high to very high changes regarding their classroom discourse performance ranging from delta values of .20 to .47.

Student engagement is not only influenced by teachers questioning behavior, but also when classroom discourse is goal-oriented and structured (quality dimension 1) (Alexander, 2005; Resnick et al., 2010) so that students can contribute in a meaningful way and can embed new information better in previous knowledge. Post TPD, the classroom discourse was more goal-oriented and structured than before. Study I investigated elaborating teacher questions, whereas in Study II it was taught in the DVC how teachers can activate their students and open their discourse through, e.g. elaborating, open-ended questions (Michaels et al., 2008; Oliveira, 2010). Teachers opened their discourse more after the TPD. Another valuable talk move, to link student answers and use these as resource (Fishman et al., 2017; Gomez Zaccarelli et al., 2018), was also implemented very well from all teachers. This was their weakest quality dimension at the beginning of the school year. This quality dimension and also the fourth, where the focus was on a warm and welcoming learning atmosphere with supportive feedback, might be the most important ones for the changes of student characteristics. Grassinger and colleagues (2018) have in this context shown that "[…] affective-motivational adaptive reactions to errors encourage action-related adaptive reactions to errors which, in turn, foster academic achievement".

The changes of classroom discourse in Study II indicate a successful TPD, as all three teachers implemented the in the DVC taught elements of dialogic teaching for improving classroom discourse. A successful implementation of so many TPD contents is seldom, as often teachers are able to implement some elements of dialogic teaching, but often they struggle then with other elements (Gomez Zaccarelli et al., 2018). Teachers need to be supported in order to implement effective practices successfully (Mercer, 2008; Pimentel & McNeill, 2013). The Dialogic Video Cycle is such a TPD program which supported teachers individually and in a learning community over one whole academic year to implement changes in their daily classroom discourse routines (see Gröschner et al., 2015), resulting in the end of the school year in a much higher quality of classroom discourse. Teachers improved all four quality dimensions, some more and some less. Several TPD programs have been facilitated and studied by their own effectiveness (van der Veen, Dobber, & van Oers, 2017; Wilkinson et al., 2016) showing that teachers only changed some elements of classroom discourse (Gomez Zaccarelli et al., 2018). Still, there were individual differences between the three teachers, not all of them changed their classroom discourse performance to the same extent and the same quality. Also, post TPD there were differences between the teachers and also with-in

each teacher between the successful implementation of the dialogic teaching elements shown as variability in the level of each quality dimension. This is also in line with previous research, which have shown that not all talk moves were implemented to the same extent (Gomez Zaccarelli et al., 2018; Michaels et al., 2008; Pehmer et al., 2015a, 2015b; Schindler et al., 2016; van de Pol et al., 2017; Vrikki et al., 2017).

9.1.2 Student Characteristics

In addition to the previously discussed findings about classroom discourse (9.1.1), this dissertation was interested in exploring how student characteristics affect classroom discourse (Study I) and how student characteristics develop (Study II) under the perspective of classroom discourse. The first model of Study I showed that student self-concept of ability was predictive of student engagement, teacher questioning behavior and teacher feedback, no predictive character of student pre-achievement for any of these patterns of verbal-teacher student interactions was found, contrary to the assumptions. Regarding self-concept of ability, this finding was in line with previous research, e.g. Jurik et al., 2014, who showed that student self-concept of ability is a crucial predictor for classroom discourse. Also, based on these findings, Study II only focused on student self-concept of ability. Study II investigated in a second step (after investigating the change of teacher's classroom discourse performance) how student self-concept of ability has changed by the end of the school year. Therefore, students were split within their class into three subgroups based on their entry level of self-concept of ability. Especially, student groups with low self-concept of ability profited from the changed classroom discourse, indicated by a positive change of their self-concept of ability. Previous studies (Blöte, 1995; Jurik et al., 2013, 2014) and also Study I of this dissertation have shown that students with higher self-concept of ability tend to be more frequently engaged. It is assumed that students with low self-concept of ability were more included in the classroom discourse after teachers participated in the DVC and so they also received more feedback which fostered their self-concept of ability. Student self-concept of ability is related to feedback in classroom discourse (Hattie & Timperley, 2007). This assumption is supported by the findings from Study I, as students with high self-concept of ability received more feedback and were also more frequently engaged. Also, students with moderate self-concept of ability profited in two classes from the changes in classroom discourse. This might also be because they received more feedback

and got more appreciation. It remains unexplained why Steve's students did not profit as much as the other two classes. Therefore, more detailed analyses of this class and the dynamics in this class might help to understand this finding. Furthermore, all three classes of Study II of this dissertation showed a slight decrease for students with high self-concept of ability. As Study I showed that the higher students' self-concept of ability was, the more they were engaged and the more feedback they received, the previously explained effects on students with low self-concept of ability might have caused the opposite effect for students with high self-concept of ability and in this sense the decrease can be explained. Meaning that students with high entry level of self-concept of ability were less involved after the TPD program and because of this they received less feedback and less appreciation causing a decrease in their self-concept of ability. The fact that Daniels students with high entry level of self-concept of ability show a slight increase shows that it is possible to change the classroom discourse practices in a positive way and to involve all students, so there is no student left behind.

9.1.3 Teacher Judgment Accuracies

Furthermore, this dissertation investigated on the one hand how teacher judgment accuracies of selected characteristics affect verbal teacher-student interactions in classroom discourse, and on the other hand how teacher judgment accuracy with regard to student self-concept of ability of certain student subgroups changed after teachers had participated in the Dialogic Video Cycle. Study I showed that it was harder for teachers to judge students' self-concept of ability than pre-achievement. This is in line with previous research, e.g. Praetorius et al. (2011) showed that teachers have difficulties in judging motivational-affective characteristics. Teachers judge the level of student pre-achievement often more accurate (Praetorius et al., 2013) as also shown in Study I (see Figure 9). Nevertheless, teacher judgment accuracy itself was not predictive of any interaction pattern of classroom discourse, but the statistical interaction of student characteristic and teacher judgment accuracy was predictive of behavior patterns in classroom discourse. Findings from Study I indicated that teacher judgment accuracy and the corresponding student characteristic are somehow related. Therefore, Study II investigated also student self-concept of ability as one of the most important student characteristics and teacher judgment accuracy of this student characteristic and how teacher's judgment accuracy changed through the TPD program more specifically

through the changed interaction. In Study I, adaptive teacher questioning behavior was found in the sense that students with lower pre-achievement received more elaborating teacher questions. Teachers tried to cognitively engaged these student to close pre-achievement gaps between students (Shavelson & Stern, 1981) as elaborating questions elicit deep thinking (Chin, 2007) and so foster learning processes. Another reason why Study II focuses on the effects of a changed classroom discourse on teacher judgment accuracy and student self-concept of ability is because Study I indicated that teachers with higher judgment accuracy of student self-concept of ability tend to provide more feedback to students with high self-concept of ability and that these students are more frequently engaged. Through implementation of elements of dialogic teaching classroom discourse performance of the three analyzed teachers improved. In a similar way, also teacher's judgment accuracies changed, resulting in higher judgment accuracies after the TPD program.

Study I underlines previous findings which have shown that teacher judgment accuracy affects teacher-student interactions. For example, already in the 1970s Brophy and Good (1970) have shown that teacher expectations affect teacher-student interactions. Brühwiler (2017) and also Sedova and Salamounova (2016) add on to this as they have shown that student participation is related to teacher judgment accuracy.

Another, common finding is from Alonzo and Kim (2018) who showed that the quality of teacher expectations was related to the discourse quality. In this sense, quality of teacher expectations can be seen as the accuracy of teacher judgments. That is why, in Study II it was assumed that teacher judgment accuracy will change as well when classroom discourse quality/performance changes. Study II investigated the development of teacher judgment accuracy of student self-concept of ability through the TPD program. Other than in Study I, Study II investigated in-depth teacher judgment accuracies of the three subgroups of students for each teacher individually. Similar changes of teacher judgment accuracy of student self-concept of ability were found. Teachers mostly judged students with low and moderate self-concept of ability more accurate after participating in the TPD program. Interestingly, also one teacher slightly improved his judgment accuracy for students with high self-concept of ability. These findings support the assumption that through the implementation of elements of dialogic teaching, a change in classroom discourse was caused which had positive effects especially on students with low self-concept of ability, resulting in more meaningful student contribution which led to more opportunities for teachers to get to know their students and in this sense to more

accurate judgments. The main source teachers can use to built their judgments on are verbal contributions of their students. Through the more intense interaction in the changed classroom discourse, teachers had greater opportunities to learn about their students' individual characteristics, thinking and learning processes. These conclusions are supported by findings from Brühwiler (2017) as he showed that student participation is significantly related to teacher judgment accuracy. These findings from Study II are, like Study I, in line with Alonzo and Kim (2018) and several other researchers (Machts et al., 2016; Sedova et al., 2016; Sedova & Salamounova, 2016) who have all shown that teacher expectations and judgments are related to teacher's classroom discourse behavior.

In general, Study I and II showed that it still remains an issue for teachers to accurately judge motivational-affective student characteristics such as student self-concept of ability (Praetorius et al., 2013; Praetorius & Südkamp, 2017; Spinath, 2005) (e.g. Figure 9), but the positive finding which is shown in Study II is that it is not a fixed construct and teachers change their judgments and can improve this competence, even though it was not addressed specifically in the TPD program.

9.2 Methodological Reflection

The two studies of this dissertation have similarities and differences in the methodology. Based on the explanations in Chapter 6.2 these will be discussed in the following as each study had its own setting, sample, and therefore certain instruments, questionnaires and video coding manuals were applied. Hence, findings have to be interpreted with regard to these circumstances and decisions. In the following these will be discussed. Furthermore, the methodological contribution of this dissertation to current research and ideas for future research will be outlined.

9.2.1 Samples

Both studies were conducted from classes of German high teaching track Gymnasium. Study I was conducted with students from 8th grade and Study II with

students from 9th and 10th grade. Students of these age groups are said to be able to express their motivational-affective characteristics (e.g. Marsh et al., 1998). When interpreting the results of the two studies, this has to be kept in mind. Furthermore, Study I and Study II have two different, independent samples and research designs (see 9.2.2), so comparisons of the two studies have to be read carefully.

In Study I only math classrooms were investigated. In Study II, there were science or math classrooms investigated. In this dissertation language art classrooms were excluded because of the domain-specitify and the different "nature" of instruction (Dweck, 1986). In future research also language arts classrooms should be investigated for the same aspects in order to get a better sense if and if so how different are language art classrooms from science or math classrooms and how accurate language art teachers judge their students and how the students subject-specific self-concept of ability affects their behavior.

9.2.2 Research Designs

Both studies were settled in different projects aiming for different research goals. Study I was based on a multi-method quantitative research design, whereas Study II was conducted as a longitudinal study and as a qualitative case study. Both studies have in common, focusing on student characteristics (Study I: pre-achievement and self-concept of ability; Study II: only self-concept of ability), as well the focus on teacher judgment accuracies (different approaches for calculation, see Chapter 9.2.3 and the overarching common ground, classroom discourse (here also different approaches to capture classroom discourse, see Chapter 9.2.3).

Both studies had their own research approach. Study I followed a more cross-sectional oriented approach whereas Study II followed a longitudinal approach. Therefore, Study I gives a broader and more widely applicable picture of how the teacher or student behavior in classroom discourse can be predicted by student characteristics and teacher judgment accuracies of student characteristics. Adding on to this, Study II shows the changes over time and as case study provides deeper insights into teacher's and their student's individual starting conditions for the TPD program and the individual uptake of the TPD program. This important resource was used to demonstrate how three teachers individually changed their classroom discourse performance and how this

affected student's self-concept of ability, again also here a differentiation based on students entry level of self-concept of ability was chosen in order to track their effects on a more fine-grained level than looking at all students in general, and also teacher's judgment accuracy of student self-concept of ability once again for each subgroup of students' entry level of self-concept of ability to track the development more precisely. This study used the advantage of case studies in order to provide a detailed and individual picture (Yin, 2014). This serves as solid basis for further research with bigger samples.

Both research approaches have its limitations which will be discussed in Chapter 9.3. Furthermore, the research design of each study has to be considered when interpreting and discussing results.

9.2.3 Instruments

The instruments used in this dissertation have to be kept in mind when interpreting the results.

9.2.3.1 Video Manuals to Assess Classroom Discourse Practices

Within each study there was a specific video coding manual used. In Study I a event sampling-based unit analysis was applied, whereas in Study II a time-sampling approach was applied. The different video codings also have to be kept in mind when interpreting the findings of the two studies of this dissertation. Study I used teacher and student talking turns as unit of analysis, Study II used 10-minute segments as unit of analysis. For Study II, time sampling was chosen over event sampling due to economic perspective and the current approaches in the field of research (Reznitskaya, Wilkinson, Oyler et al., 2016). Furthermore, Study I showed that scarce events were analyzed resulting in low explained variance, therefore time sampling also has an advantage as these are high-inference video codings (Begrich, Fauth, Kunter, & Klieme, 2017). Further codings of Study I were then coded based on a detailed coding manual to student engagement, elaborating teacher questions and supportive teacher feedback. The quality of classroom discourse in Study II was assessed using a theory-based coding manual with four quality dimensions which captured the most important components of a productive classroom discourse. Codings from Study I are implicitly found in the quality dimensions. In this sense, the findings of the two studies can be compared (see Chapter 9.1) when keeping these facts in mind.

9.2.3.2 Student Characteristics

Student pre-achievement was used as cognitive characteristic in Study I. It was measured as grades from the previous academic year. The use of grades as reliable measurement of a student characteristic is often criticized as grades are already based on teachers' judgments. "Grading refers to the symbols assigned to individual pieces of student work or to composite measures of student performance on student report cards" (Brookhart et al., 2016). In this sense, grading contains cognitive and non-cognitive factors. Critique on the use of grades for measuring pre-achievement is reasonable and therefore and because of the relevance of student self-concept of ability for verbal interactions Study II investigates only self-concept of ability. Still, the use of pre-achievement measured as grades in Study I is appropriate as these are high predictors for student learning success (e.g. Hattie, 2008) and as such they were applicable for the fine grained investigation of classroom discourse in Study I. For assessing student self-concept of ability in the two studies there were used different scales. Study I used a scale from PISA (Hertel et al., 2014) and Study II applied a scale from Marsh et al. (Marsh et al., 2005), both on a four-point Likert scale. Using student self-report assessed through the mentioned scales seems to be appropriate in this dissertation and is a common practice as student characteristics in relation to teachers' judgments were investigated. Furthermore, students of the investigated age (8^{th} to 10^{th} grade) are able to evaluate their characteristics in a reliable way (Marsh et al., 2005). Furthermore, Study II focused more on student subgroups within each class to track the changes of these subgroups in detail, whereas Study I investigated 18 classes as a whole sample to get the status quo.

9.2.3.3 Teacher Judgment Accuracy

Regarding teacher judgments, Study II was expanded with regard to the scale level teacher judgments were assessed. In Study I, teachers judged their students for each characteristic on a three-point Likert scale, whereas teachers in Study II judged their students on a five-point Likert scale. The expand of the Likert scale made sense and was good as teachers had the chance to differentiate more within their class. In both studies, teacher judgment data and student characteristic data had to be transformed into a suitable scale level. For future research, it would be recommended to have the same scale level for teacher and student data when aiming for calculating accuracies or any

comparisons between teacher and student scores, to exclude any errors which might be caused through different scale levels.

This dissertation explored and applied two different approaches to calculate teacher judgment accuracy. Study I used the most commonly applied method, to calculate teacher judgment accuracy as Spearman correlation of teacher judgment and student characteristic (Machts et al., 2016; Praetorius et al., 2011; Praetorius et al., 2013; Spinath, 2005; Südkamp et al., 2012). It was decided to report results based on this measurement approach to keep the results comparable to other research in this field and as there was no such individual level needed. But it was already explored in Study I to calculate teacher judgment accuracy on an individual level as difference between teacher judgment and student characteristic (Thiede et al., 2018). As there were no differences for Study I, it was decided to stay with the more wide spread measurement of teacher judgment accuracy (correlation). In Study II, again both approaches were explored, it was shown that the individual measurement is more sensible when analyzing in-depth cases and splitting into subgroups and do not look at the overall picture. Therefore, the individual teacher judgment accuracy (difference) was suitable for Study II as it is on an individual student level and Study II investigated in-depth teachers' individual changes and also student subgroups within each class. Calculating teacher judgment accuracy on an individual level adds on to the movement in the research field of teacher diagnostic competences (Praetorius, Koch, Scheunpflug, Zeinz, & Dresel, 2017; Südkamp & Praetorius, 2017) and serves as basis for further research with regard to teacher individual changes, also maybe with a bigger sample and not only in case studies.

9.3 Limitations and Directions for Future Research

Several limitations have to be considered within this dissertation. Each of this dissertations studies' limitations have already been discussed in Chapter 7.7.3 for Study I and Chapter 8.7.4 for Study II. In the following the limitations will be brought together and summarized as limitations of the whole dissertation. Furthermore, this dissertation also raised new research questions and points out some directions for future research which will be outlined also within this Chapter.

Regarding Study II, all teachers participated voluntarily in the TPD program. In this sense this is a limitation as a so-called positive selection biases the findings because only motivated teachers who were willing to change their classroom discourse practices participated and sacrifice additional time. Whereas Study I chose schools who had to participate in this study, if these would have participated in the TPD program then maybe the effects of the TPD program might look different. In order to decline such hypotheses, it would be interesting to investigate teachers who do not participate voluntarily. But from the practical perspective this is impossible as teacher have to spend much time and energy in addition to their daily teaching into the TPD program. One approach could be that facilitators from universities go directly to schools and offer a TPD program for all teachers or at least one teacher of each subject. But again from a research perspective, this would bring up the next problem that then all teachers are nested in one school.

Assessing student pre-achievement as grades is critically discussed. The fact that the distribution was approximately normal within this high-achieving group even though it was German Gymnasium brings up the discussion about the sense of German or Bavarian School System where students are separated according to their grades after elementary school into three types of school (Gymnasium, Realschule, Mittelschule) (Drewek, 2013; Hurrelmann, 2013). PISA has also shown that students with regard to their mathematical and their reading competence have a wide range and that the student from different school types are slightly different but not that on Gymnasium only high achieving students are (Klieme, 2010; Sälzer & Reiss, 2016). Also, student self-concept of ability showed wide ranges (see Figure 8). In Study II also high variation within each class was found.

Study I has as another limitation that student characteristics were measured at the beginning of the school year and then linked to classroom discourse behavior measured about three to four months after school began. In Study II classroom discourse behavior and questionnaire data were measured at the same time, once at the beginning and once at the end of the school year. This leads right to the next limitation of both studies of this dissertation: only one lesson unit was videotaped for each measuring point. Even tough research has shown that quite typical and frequently occurring teaching routines are represented which are likely to occur the same way in other lessons throughout the school year (Praetorius et al., 2014; Seidel & Prenzel, 2006), it might be interesting to video-tape more than one lesson. The exact amount of how many lesson units are required for representative, valid and reliable behavioral findings is not yet known and needs more

research (Praetorius et al., 2014). Furthermore, with regard to teacher's professional developments additional video analysis throughout the school year might yield additional information on the individual trajectories of classroom discourse practice changes.

Even though two different video manuals were used for coding the investigated lessons of this dissertation, they have both their limitations. In both studies, the observed teacher behavior could be to some extent interrelate with student behavior. Study I focused also only on a very specific type of questions and also regarding teacher feedback which also indicated that not every student contribution received supportive feedback, therefore also evaluative feedback as additional feedback could be investigated in future research to capture if there is any teacher response to student contribution and then to what quality. In Study II, to some extent the call for investigating feedback further from Study I was heard in the sense that feedback can be seen as part of quality dimension 4 (supportive and scaffolding) and so it was investigated. The other limitation of the applied video coding manual is that it based on a high-inference rating system meaning that high level of conclusions on the part of the observers are implicated (Lotz et al., 2013). As implication for future research could be that high- and low-inference video codings are combined. Furthermore, with regard to Study II analyses of student behavior would be interesting to find corresponding behaviors and to prove the assumptions why and how student self-concept of ability changed through the TPD program.

Another problem this dissertation had to face in both studies was the possible loose of information through building percentile groups in order to match teacher and student data. In Study I, student information was lost in the way that the exact performance of each student had to be recoded to be matched with the three-point Likert scale of teacher judgments. In Study II, teacher judgment information was lost in the way that the exact teacher judgment of each student had to be recoded to be matched with the four-point Likert scale of student data. To avoid this in future studies, scale levels of scales which later will be combined or used to e.g. calculate teacher judgment accuracy no matter if as correlation or difference, should be the same from the beginning on.

With regard to teachers' judgment and teacher judgment accuracies it has to be mentioned that neither in Study I nor in Study II it was in the focus of the projects, though these were investigated. It is difficult to argue that accurate teacher judgments might result in specific teaching acts, or the other way round, that specific teaching acts influence teacher judgment accuracies per se. But the two studies gave hints that such relationships are given. And also other researchers, showed that teacher expectations and

judgments were to an extent related to the teacher's discourse behavior (Machts et al., 2016; Sedova et al., 2016; Sedova & Salamounova, 2016).

With regard to Study II, it is not known/controlled if teachers were maybe triggered for their judgments as they had to fill out the questionnaires at the beginning, in the middle and at the end of the school year. It could be that teachers were more aware of their judgments because of that. Therefore, future research should control for this by either pick it up as topic in the TPD program or use further data sources, e.g. interview data to control therefore.

In general, TPD programs not only focusing on improving classroom discourse should be aware of the hidden effects which might be caused through the TPD program. Focusing in a TPD program on one aspect, in this project classroom discourse, might also effect teacher judgment accuracy as shown in Study II or student self-concept of ability. For conceptualizing TPD programs, researchers should consider all kinds of effects which it might have, therefore an orientation was given by Desimone (2009) who suggested a chain of effects caused by TPD program, but in this sense and I think that is also something what this dissertation has shown that there are often more complex interactions, and maybe it is not a linear chain, but more a circle where also in between relations/effects might have to be considered (Schindler et al., submitted).

Study II gives some more implications for future research as it was a case study. Case studies can serve as basis to get deep insights into a complex system (Yin, 2014), such as in Study II teachers individual trajectories. This serves as basis for future research investigations with a bigger sample. Furthermore, case studies can show what can be accomplished. In Study II of this dissertation it was demonstrated that teachers can change their teaching routines and that this also seems to have effects on student characteristics such as self-concept of ability and on teacher judgment accuracies. But case studies are different to generalize and in this sense, they often have only limited significance (Yin, 2014). So, this dissertation delivers a basis for further research with bigger samples investigating the development of teacher judgment accuracies in relation to changes in classroom discourse practices through a TPD program.

9.4 Educational Relevance and Implications

This dissertation adds on to current research and gives several implications for practice. Findings from both studies of this dissertation are highly relevant for teacher education, teacher professional development and daily teaching practices. These will be explained and discussed in the following chapter.

9.4.1 Educational Relevance and Implications for Teacher Education and Teacher Professional Developments

Teacher professional development and teacher education have been called to be structured around *high-leverage practices* (e.g. McDonald et al., 2013). In this chapter it is discussed how the findings of this dissertation can be implemented in teacher education and teacher professional development on the one hand to close the gap between theory and practice and on the other hand how and what teachers and student teachers should be taught in order to improve their teaching.

9.4.1.1 Classroom Discourse Practices

Classroom discourse is still the main teaching method (e.g. Hiebert et al., 2003). This dissertation has shown that specific teaching patterns of classroom discourse can be predicted by student characteristics and teacher judgment accuracy (Study I) and that classroom discourse can be changed (Study II). With regard to teacher education it should be taught that teaching is a life-long learning process and classroom discourse as the main teaching method which has to be optimized and adapted to students' individual needs.

The importance of classroom discourse has been in focus of research for many years now (Alexander, 2005; Mercer & Dawes, 2014). Still, there has been progress and changes have been made. Also, the focus when investigating classroom discourse has developed. For example, O'Connor and colleagues (2017) brought up the discussion about the silent and the vocal student and who learns better. In their study, they have shown that with regard to their learning outcomes it did not matter if students were verbal active participating or passive participators as active listeners. Teachers need to learn on the one side to encourage students to verbally contribute to ongoing classroom discourse and on the other hand also to actively listen to others' contributions. The active listening

component is a basis for linking students answers which was brought up from the Stanford group around Jonathan Osborne. The focus of this talk move is to relate student contributions for achieving a common understanding (Gomez Zaccarelli et al., 2018). Findings of this dissertation underline these previous findings and highlight the importance to improve classroom discourse. Classroom discourse has many different facets, besides student engagement, also teacher questioning behavior needs to be in the focus of teacher education and also teacher professional development. Study I has shown that elaborating questions, as questions which intend students to explain or elaborate facts that are not yet known in this form that requires deeper thinking, were scarce. Wilkinson, Reznitskaya and colleagues (2016) for example go that far that they claim that sometimes only one central question is enough or should be enough to circle classroom discourse around that question. This might apply more for language arts classrooms, but still for some parts of lessons or for some topics also in science classrooms it should be explored. The power of feedback have also Hattie and Timperley already stated to investigate and postulate in 2007. Feedback is only useful if it contains information for the students further learning or/and thinking process (Hattie & Timperley, 2007). Study I and II also showed that this is an important component of classroom discourse which teachers and student teachers have to be prepared for. A whole TPD program could and maybe also should be filled with the importance of feedback and how to provide feedback adequately. In teacher education, this could be done right before teacher students go into schools to get experience in teaching then they should be sensitized for good feedback and can explore how to provide feedback, afterwards a reflection should be made, based on the same principle as in the Dialogic Video Cycle.

The often described learning atmosphere as one central criterion for good teaching practices (Meyer, 2016) should be also included in teacher education and teacher professional development programs. An active and fearless learning atmosphere is very important for student learning success (Grassinger et al., 2018) . How to achieve this needs to be taught. Study II showed that this can be changed through professional development (see positive changes in quality dimension 4 of each teacher (Figure 11a, 12a, 13a). If students are not afraid to contribute to the discourse, teachers cannot learn about their possible misconceptions or struggles, only then teachers can support their students in their further learning process.

Another often mentioned criterion for good teaching practices is goal-orientation, also in classroom discourse this is a central quality criterion for productive classroom discourse and therefore has to be taught in teacher education and teacher professional development programs. Goal-orientation is a fundamental basis (Resnick et al., 2010) as students need to know what the classroom discourse is about or where it should go, otherwise no meaningful contributions will be possible. Positive changes of quality dimension 1 and 2 (see Figures 11a, 12a, 13a) point out the relevance and show that changes with regard to these elements are possible if teachers are advised good how to implement changes into their classrooms.

To sum up, findings of this dissertation's studies call for supporting teachers and student teachers in acquiring knowledge about classroom discourse in its diverse manners and how to improve this main teaching method throughout their professional career.

9.4.1.2 Student Characteristics

Teaching processes are to a large extent determined by verbal interactions between teachers and their students (Michaels & O'Connor, 2012; Walshaw & Anthony, 2008). In this context, the individual characteristics of the main characters play an important role. As this dissertation focused on student characteristics the relevance of these for teachers will be outline in the following. Teachers and student teachers should learn about the importance of student characteristics as every student is individual (Huber, 2017). Student pre-achievement is an important characteristic for students further learning success and as it can be regarded as an outcome of instruction. Findings from Study I of this dissertation have shown that especially student self-concept of ability is important for classroom discourse and Study Ii has underlined this importance. Therefore, teacher education and teacher professional development programs should foster the importance of e.g. self-concept of ability. This dissertation added to previous studies (Jurik et al., 2014; Marsh & Martin, 2011) as it was shown that self-concept of ability is especially important in the context of classroom discourse. Teachers and student teachers should learn how important student self-concept of ability is as students with high self-concept of ability are more frequently engaged (e.g. Jurik et al., 2014). In this sense, the appearance of the so called 'Matthew'-effect (see Chapter 1.1) should be prevented in the way that the gap between students with higher or lower characteristics

does not grow or is being closed. Findings from Study II indicate that through the interaction with students with lower self-concept of ability their self-concept of ability increased and the self-concept of ability of students with higher self-concept of ability decreased (see for example Case Lisa Figure 11b).

9.4.1.3 Teacher Judgment Accuracy

In general, teachers should learn about the importance of the students' individual characteristics. Only if they know about their importance, they can be more aware of these and this brings up the next implication. This dissertation has once more brought evidence that teachers have difficulties in judging motivational-affective student characteristics accurately. Teachers and student teachers need to learn first that students are individuals and diverse because the problem is that teachers often only see the "overall strong" or the "struggling" students (Huber et al., 2015; Huber, 2017) and second teachers and student teachers need to learn how to accurately assess their students characteristic (Machts et al., 2016). Therefore, teachers and student teachers need to learn about the different cues which seem to be the reason for the difficulties in judging motivational-affective characteristics accurately (Pielmeier, Huber et al., 2018). Diagnosing student characteristics accurately is a professional demand and necessary for providing instruction on an individual level, in other words in order to teach adaptively (Corno, 2008).

Accurate teacher judgments are the most common way to measure teacher diagnostic competences. Teacher diagnostic competence is rarely taught in teacher education and teacher professional development even though it is crucial for teaching. Student teachers should be prepared more with knowledge on diagnostic competence and acquire these during their studies. TPD programs should support in-service teachers to improve their diagnostic competences and teacher education should teach student teachers how to acquire good diagnostic competences. In the last few years, there have been established research groups and networks, e.g. NeDiKo (Netzwerk – Diagnostische Kompetenz) and COSIMA research group, focusing on teacher diagnostic competence. For example, the COSIMA research group is one example who are investigating how simulation-based learning environments can be designed and used to foster and improve teacher diagnostic competence in teacher and medicine students' studies, e.g. research subproject VisitMath. Within this project it is explored how simulation-based learning

environments can be used in teacher education to promote diagnostic skills of mathematics teacher students. Such projects and research groups can develop and establish evidence-based standards for teacher diagnostic competence in teacher education.

As shown in Study II of this dissertation, effectively designed teacher professional development programs can change teaching routines and therefore remain indispensable (Schindler et al., submitted) for ongoing and life-long learning for teachers. Furthermore, this dissertation has shown that teacher professional development programs might cause also other changes, e.g. in this case improvement of teachers' judgment accuracies. Therefore, future TPD programs should be aware of the possible silent outcomes.

9.4.2 Educational Relevance and Implications for Daily Teaching Practices

Previously, the educational relevance of this dissertation's findings for teacher education and teacher professional development more in a long-term and broader sense have been explained. Now, the focus is on what this dissertation delivers for teachers' daily teaching practices.

Findings of this dissertation are highly relevant for teachers in their daily teaching lives as classroom discourse is the main teaching method (Hiebert et al., 2003). The finding of Study I, that student self-concept of ability is predictive of student engagement, elaborating teacher questions and supportive teacher feedback should show teachers the high relevance of student self-concept of ability. Keeping this in mind in daily teaching, teachers should plan and implement their instruction always towards and based on their students' individual needs.

The findings from Study II show teachers that student self-concept of ability is a sensitive student characteristic that can be changed through and are formed by (changed) interactions in classroom discourse. For daily teaching practices teachers should learn about the little things that can cause changes, i.e. through elements of dialogic teaching (Pielmeier et al., submitted; Schindler et al., submitted). But as changes can go in either direction, positive or negative, it is recommended that teachers participate in professional development programs because there is guidance when teachers start to make changes to

their daily teaching routines, if the program follows the core features of successful professional development (Desimone, 2009).

For student characteristics, such as self-concept of ability, feedback plays an especially important role (e.g. Pielmeier, Huber et al., 2018). Only students who are verbally active can receive supportive feedback. In the long run, this can lead to a higher self-concept of ability. If teachers try to pay attention to more and a variety of students and involve more and various students in their daily classroom discourse and then providing each student supportive, not evaluative feedback, teachers could change their classroom discourse and their students self-concept of ability, especially that from students with low self-concept of ability (Pielmeier et al., submitted). It is a lot teachers have to take care of and the role of teachers is continuously growing (O'Connor et al., 2017). But the in Study II introduced elements of dialogic teaching, represented in the quality dimensions of classroom discourse, can be implemented in each classroom discourse. To control therefore, teachers could use peer-to-peer feedback (Helmke et al., 2018), in the way of classroom observations by colleagues. Therefore, the video coding manual from Study II could also be used or serve as basis for structured observation and constructive feedback. Hence, feedback from collegiate classroom observations have been shown to have positive effects e.g. on teachers motivation, self-efficacy and also self-regulating/self-reflecting processes (Funk, 2016), this could be implemented easily in daily teaching practices.

Classroom discourse, especially teachers' behavior, is also influenced by teachers' expectations and judgments (Sedova & Salamounova, 2016). As it was shown that judgment quality and discourse quality are closely related together (Alonzo & Kim, 2018), Study II adds on to these findings showing that changes in discourse quality lead also to changes in teachers' judgment accuracies. Teachers can learn from this for their daily teaching that first of all it is important that teachers are open-minded to change and revise their judgments of students. Teachers should be aware that they often misjudge their students especially regarding motivational-affective characteristics (Praetorius et al., 2011; Praetorius et al., 2013). This is especially important to adapt teaching to meet students' individual needs. For adaptive teaching, teachers need to judge their students accurately, otherwise their adapted instruction is based on an inaccurate diagnosis. In order to avoid such, teachers could also use questionnaires, as the ones applied in this dissertation's studies, to assess their students' individual characteristics. Comparisons with teachers' own judgments could lead to eye-opening experiences.

9.5 Overall Conclusion

This dissertation investigated the interplay of teacher judgment accuracies and student characteristics in classroom discourse. Classroom discourse can be investigated from different angles. This dissertation focused on the following two: first, on a specific level where the typical occurring IRF patterns were found. Second, teachers' classroom discourse performance was measured on a more general level with quality dimensions of productive classroom discourse. This highlighted that classroom discourse can be successfully changed through a video-based professional development program.

Furthermore, this dissertation underlines the importance of student self-concept of ability which is often discussed in the field of educational research. This dissertation did so not only for student engagement, but also for classroom discourse practices in general.

Indications that teacher judgments can serve as basis for adaptive teaching behavior were found and teachers' difficulties in judging motivational-affective characteristics accurately were once again highlighted. It was also shown that these are not fixed patterns, so judgment accuracy can improve if there are more opportunities in classroom discourse for teachers to get to know their students better. Furthermore, teacher judgment accuracy improved, even though it was not topic of the program, indicating hidden effects of TPD programs. These possible silent hidden effects should be considered more systematically in future research on TPD. Overall, this dissertation has shown that teachers change their daily routines if guided properly.

Case studies can serve as basis for further research, provide deep insights into teachers individual learning trajectories within a TPD program and also show other components influenced by such changes in classroom discourse behavior.

All in all, this dissertation shows that focusing on individual learning and characteristics in educational research and practice can lead to successful learning on teacher and student side and teaching can be improved.

References

Aguiar, O. G., Mortimer, E. F., & Scott, Phil. (2010). Learning from and responding to students' questions: The authoritative and dialogic tension. *Journal of Research in Science Teaching*, *10*, n/a-n/a. https://doi.org/10.1002/tea.20315

Alexander, R. J. (2005). *Towards dialogic teaching: Rethinking classroom talk* (2nd edition). Cambridge: Dialogos.

Alles, M., Seidel, T., & Gröschner, A. (2018). Establishing a positive learning atmosphere and conversation culture in the context of a video-based teacher learning community. *Professional Development in Education*, *11*(1), 1–14. https://doi.org/10.1080/19415257.2018.1430049

Alonzo, A. C., & Kim, J. (2018). Affordances of video-based professional development for supporting physics teachers' judgments about evidence of student thinking. *Teaching and Teacher Education*. (76), 283–297. https://doi.org/10.1016/j.tate.2017.12.008

Anders, Y., Kunter, M., Brunner, M., Krauss, S., & Baumert, Jürgen. (2010). Mathematics Teachers' Diagnostic Skills and Their Impact on Students' Achievements. *Psychologie in Erziehung und Unterricht*, *57*(3), 175–193. https://doi.org/10.2378/peu2010.art13d

Ball, D. L., & Forzani, F. M. (2009). The Work of Teaching and the Challenge for Teacher Education. *Journal of Teacher Education*, *60*(5), 497–511. https://doi.org/10.1177/0022487109348479

Begrich, L., Fauth, B., Kunter, M., & Klieme, E. (2017). How informative are first impressions?: The thin-slices technique as a method for video-based assessment of instruction. *Zeitschrift für Erziehungswissenschaft.* Advance online publication. https://doi.org/10.1007/s11618-017-0730-x

Bickel, R. (Ed.). (2007). *Methodology in the social sciences*. *Multilevel analysis for applied research: It's just regression!* New York: Guilford Press.

Bleicher, R. E., Tobin, K. G., & McRobbie, C. J. (2003). Opportunities to Talk Science in a High School Chemistry Classroom. *Research in Science Education*, *33*, 319–339.

Blomberg, G., Sherin, Miriam G., Renkl, A., Glogger, I., & Seidel, T. (2014). Understanding video as a tool for teacher education: Investigating instructional strategies to promote reflection. *Instructional Science*, *42*(3), 443–463. https://doi.org/10.1007/s11251-013-9281-6

Blöte, A. W. (1995). Students' self-concept in relation to perceived differential teacher treatment. *Learning and Instruction*, *5*(3), 221–236. https://doi.org/10.1016/0959-4752(95)00012-R

Blumenfeld, P. C., Pintrich, P. R., Meece, J., & Wessels, K. (1982). The Formation and Role of Self Perceptions of Ability in Elementary Classrooms. *The Elementary School Journal*, *82*(5), 401–420. https://doi.org/10.1086/461278

Blut, M. (2008). *Der Einfluss von Wechselkosten auf die Kundenbindung: Verhaltenstheoretische Fundierung und empirische Analyse [The influence of exchange costs on costumer loyalty: behavioral theoretical foundation and empirical analysis]*. *Gabler Edition Wissenschaft*. Wiesbaden: Betriebswirtschaftlicher Verlag Dr. Thomas Gabler / GWV Fachverlage GmbH Wiesbaden. https://doi.org/10.1007/978-3-8349-9795-1

Borko, H. (2004). Professional Development and Teacher Learning: Mapping the Terrain. *Educational Researcher*, *33*(8), 3–15. https://doi.org/10.3102/0013189X033008003

Borko, H., Jacobs, Jennifer, Eiteljorg, E., & Pittman, Mary Ellen. (2008). Video as a tool for fostering productive discussions in mathematics professional development. *Teaching and Teacher Education*, *24*(2), 417–436. https://doi.org/10.1016/j.tate.2006.11.012

Borko, H., Jacobs, Jennifer, Seago, N., & Mangram, C. (2014). Facilitating Video-Based Professional Development: Planning and Orchestrating Productive Discussions. In Y. Li, E. A. Silver, & S. Li (Eds.), *Advances in Mathematics Education. Transforming Mathematics Instruction: Multiple Approaches and Practices* (pp. 259–281). Cham, s.l.: Springer International Publishing. https://doi.org/10.1007/978-3-319-04993-9_16

Borko, H., Roberts, S. A., & Shavelson, Richard. (2008). Teachers' Decision Making: From Alan J. Bishop to Today. In P. Clarkson & N. Presmeg (Eds.), *Critical Issues in Mathematics Education: Major Contributions of Alan Bishop* (1st ed., pp. 37–67). Boston, MA: Springer. https://doi.org/10.1007/978-0-387-09673-5_4

Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J., Smith, L., . . . Welsh, M. E. (2016). A Century of Grading Research: Meaning and Value in the Most Common Educational Measure. *Review of Educational Research, 86*(4), 803–848. https://doi.org/10.3102/0034654316672069

Brophy, J. E., & Good, T. L. (1970). Teachers' communication of differential expectations for children's classroom performance: Some behavioral data. *Journal of Educational Psychology, 61*(5), 365–374. https://doi.org/10.1037/h0029908

Brown, G. A., & Wragg, E. C. (2001). *Questioning in the Secondary School* (2nd edition). Hoboken: Taylor and Francis.

Brühwiler, C. (2017). 5.3 Diagnostische und didaktische Kompetenz als Kern adaptiver Lehrkompetenz. In A. Südkamp & A.-K. Praetorius (Eds.), *Pädagogische Psychologie und Entwicklungspsychologie. Diagnostische Kompetenz von Lehrkräften: Theoretische und methodische Weiterentwicklungen [Diagnostic Competence of Teachers: Theoretical and Methodical Evolution. Educational psychology and developmental psychology]* (pp. 123–134). Münster, New York: Waxmann.

Brunswik, E. (1952). *The conceptual framework of psychology*. Oxford: University of Chicago Press.

Buzzelli, C., & Johnston, B. (2001). Authority, power, and morality in classroom discourse. *Teaching and Teacher Education, 17*, 873–884.

Cazden, C. B. (2001). *Classroom discourse: The language of teaching and learning* (2nd edition). Portsmouth, NH: Heinemann.

Chafi, M. E., & Elkhouzai, E. (2016). The Use of Feedback in Classroom Interaction in Moroccan Primary School. *European Scientific Journal, ESJ, 12*(4), 281. https://doi.org/10.19044/esj.2016.v12n4p281

Chi, M. T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science, 1*(1), 73–105. https://doi.org/10.1111/j.1756-8765.2008.01005.x

Chin, C. (2006). Classroom Interaction in Science: Teacher questioning and feedback to students' responses. *International Journal of Science Education, 28*(11), 1315–1346. https://doi.org/10.1080/09500690600621100

Chin, C. (2007). Teacher questioning in science classrooms: Approaches that stimulate productive thinking. *Journal of Research in Science Teaching, 44*(6), 815–843. https://doi.org/10.1002/tea.20171

Cohen, J. (2010). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd edition). New York, NY: Routledge.

Coles, A. (2013). Using video for professional development: The role of the discussion facilitator. *Journal of Mathematics Teacher Education*, *16*(3), 165–184. https://doi.org/10.1007/s10857-012-9225-0

Corno, L. (2008). On Teaching Adaptively. *Educational Psychologist*, *43*(3), 161–173. https://doi.org/10.1080/00461520802178466

Cortina, K. S., Miller, K. F., McKenzie, R., & Epstein, A. (2015). Where Low and High Inference Data Converge: Validation of CLASS Assessment of Mathematics Instruction Using Mobile Eye Tracking with Expert and Novice Teachers. *International Journal of Science and Mathematics Education*, *13*(2), 389–403. https://doi.org/10.1007/s10763-014-9610-5

Dai, D. Y., & Sternberg, R. J. (Eds.). (2004). *The educational psychology series*. *Motivation, emotion, and cognition: Integrative Perspectives on Intellectual Functioning and Development*. Mahwah, N.J: Lawrence Erlbaum Associates.

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, *125*(6), 627–668. https://doi.org/10.1037/0033-2909.125.6.627

Demaray, M. K., & Elliott, S. N. (1998a). Teachers' Judgments of Students' Academic Functioning: A Comparison of Actual and Predicted Performances. *School Psychology Quarterly*, *13*(1), 8–24.

Demaray, M. K., & Elliott, S. N. (1998b). Teachers' judgments of students' academic functioning: A comparison of actual and predicted performances. *School Psychology Quarterly*, *13*(1), 8–24.

Desimone, L. M. (2009). Improving Impact Studies of Teachers' Professional Development: Toward Better Conceptualizations and Measures. *Educational Researcher*, *38*(3), 181–199. https://doi.org/10.3102/0013189X08331140

Dochy, F., Rijdt, C. de, & Dyck, W. (2002). Cognitive Prerequisites and Learning. *Active Learning in Higher Education*, *3*(3), 265–284. https://doi.org/10.1177/1469787402003003006

Drewek, P. (2013). Das dreigliedrige Schulsystem im Kontext der politischen Umbrüche und des demographischen Wandels im 20. Jahrhundert. *Zeitschrift für Pädagogik*, *59*(4), 508–525.

Duffy, G., & Roebler, L. (1986). Constraints on Teacher Change. *Journal of Teacher Education*, *37*(1), 55–58. https://doi.org/10.1177/002248718603700112

Dweck, C. S. (1986). Motivational Processes Affecting Learning. *American Psychologist*, *41*(10), 1040–1048.

Feinberg, A. B., & Shapiro, E. S. (2003). Accuracy of teacher judgments in predicting oral reading fluency. *School Psychology Quarterly*, *18*(1), 52–65. https://doi.org/10.1521/scpq.18.1.52.20876

Fiedler, K., Walther, E., Freytag, P., & Plessner, H. (2002). Judgment Biases in a Simulated Classroom—A Cognitive-Environmental Approach. *Organizational Behavior and Human Decision Processes*, *88*(1), 527–561. https://doi.org/10.1006/obhd.2001.2981

Fishman, E. J., Borko, H., Osborne, Jonathan, Gomez, F., Rafanelli, S., Reigh, E., . . . Berson, E. (2017). A Practice-Based Professional Development Program to Support Scientific Argumentation From Evidence in the Elementary Classroom. *Journal of Science Teacher Education*, *28*(3), 222–249. https://doi.org/10.1080/1046560X.2017.1302727

Funk, C. M. (2016). *Kollegiales Feedback aus der Perspektive von Lehrpersonen: Zusammenhang von beruflicher Zielorientierung und Bewertung des Feedbackkonzepts*. Wiesbaden: Springer VS. Retrieved from http://gbv.eblib.com/patron/FullRecord.aspx?p=4526702

Furtak, E. M., & Kunter, M. (2012). Effects of Autonomy-Supportive Teaching on Student Learning and Motivation. *The Journal of Experimental Education*, *80*(3), 284–316. https://doi.org/10.1080/00220973.2011.573019

Gan, M. J. S., & Hattie, John. (2014). Prompting secondary students' use of criteria, feedback specificity and feedback levels during an investigative task. *Instructional Science*, *42*(6), 861–878. https://doi.org/10.1007/s11251-014-9319-4

Gaudin, C., & Chaliès, S. (2015). Video viewing in teacher education and professional development: A literature review. *Educational Research Review*, *16*, 41–67. https://doi.org/10.1016/j.edurev.2015.06.001

Glogger-Frey, I., Deutscher, M., & Renkl, A. (2018). Student teachers' prior knowledge as prerequisite to learn how to assess pupils' learning strategies. *Teaching and Teacher Education*, *76*, 227–241. https://doi.org/10.1016/j.tate.2018.01.012

Gomez Zaccarelli, F., Schindler, A.-K., Borko, H., & Osborne, Jonathan. (2018). Learning from professional development: A case study of the challenges of enacting

productive science discourse in the classroom. *Professional Development in Education*, *9*(2), 1–17. https://doi.org/10.1080/19415257.2017.1423368

Grassinger, R., Scheunpflug, A., Zeinz, H., & Dresel, M. (2018). Smart is who makes lots of errors?: The relevance of adaptive reactions to errors and a positive error climate for academic achievement. *High Ability Studies*, *29*(1), 37–49. https://doi.org/10.1080/13598139.2018.1459294

Greene, W. H. (2012). *Econometric analysis* (7th edition, international edition). *Pearson series in economics*. Boston, London: Pearson.

Gröschner, A., Schindler, A.-K., Holzberger, D., Alles, M., & Seidel, T. (2018). How systematic video reflection in teacher professional development regarding classroom discourse contributes to teacher and student self-efficacy. *International Journal of Educational Research.* Advance online publication. https://doi.org/10.1016/j.ijer.2018.02.003

Gröschner, A., Seidel, T., Kiemer, K., & Pehmer, A.-K. (2015). Through the lens of teacher professional development components: The 'Dialogic Video Cycle' as an innovative program to foster classroom dialogue. *Professional Development in Education*, *41*(4), 729–756. https://doi.org/10.1080/19415257.2014.939692

Gröschner, A., Seidel, T., Pehmer, A.-K., & Kiemer, K. (2014). Facilitating collaborative teacher learning: The role of "mindfulness" in video-based teacher professional development programs. *Gruppendynamik und Organisationsberatung*, *45*(3), 273–290. https://doi.org/10.1007/s11612-014-0248-0

Grossman, P., Hammerness, K., & McDonald, M. (2009). Redefining teaching, re-imagining teacher education. *Teachers and Teaching*, *15*(2), 273–289. https://doi.org/10.1080/13540600902875340

Hammer, S., Reiss, K., Lehner, m. C., Heine, J.-H., Sälzer, C., & Heinze, A. (2016). Mathematische Kompetenz in PISA 2015: Ergebnisse, Veränderungen und Perspektiven [Mathematical competence in PISA 2015: results, changes and perspectives]. In K. Reiss, C. Sälzer, A. Schiepe-Tiska, E. Klieme, & O. Köller (Eds.), *PISA 2015: Eine Studie zwischen Kontinuität und Innovation [PISA 2015: A Study between Continuity and Innovation]* (pp. 219–247). Münster, New York: Waxmann.

Hattie, John A. C. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.

Hattie, John A. C., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

Helmke, A. (2009). *Unterrichtsqualität und Lehrerprofessionalität: Diagnose, Evaluation und Verbesserung des Unterrichts ; Franz Emanuel Weinert gewidmet* (2. aktualisierte Aufl.): Seelze-Velber; Klett/Kallmeyer.

Helmke, A., Helmke, T., Lenske, G., Pham, G., Praetorius, A.-K., Schrader, F.-W., & Ade-Thurow, M. (2018). Evidenzbasierte Methoden der Unterrichtsdiagnostik und -entwicklung. Retrieved from www.unterrichtsdiagnostik.info

Hennessy, S., Rojas-Drummond, S., Higham, Rupert, Márquez, A. M., Maine, F., Ríos, R. M., . . . Barrera, M. J. (2016). Developing a coding scheme for analysing classroom dialogue across educational contexts. *Learning, Culture and Social Interaction*, *9*, 16–44. https://doi.org/10.1016/j.lcsi.2015.12.001

Herppich, S., Praetorius, A.-K., Hetmanek, A., Glogger-Frey, I., Ufer, S., Leutner, D., . . . Südkamp, A. (2017). Teachers' assessment competence: Integrating knowledge-, process-, and product-oriented approaches into a competence-oriented conceptual model. *Teaching and Teacher Education.* Advance online publication. https://doi.org/10.1016/j.tate.2017.12.001

Hertel, S., Hochweber, J., Mildner, D., Steinert, B., & Jude, N. (2014). *PISA 2009 Skalenhandbuch [PISA 2009 Scale Manual]*. Münster: Waxmann.

Hiebert, J., Gallimore, R., Givvin, K. B., Hollingsworth, H., Jacobs, Jennifer, Chui, A. M.-Y., . . . Stigler, J. (2003). Teaching Mathematics in Seven Countries: Results From the TIMSS 1999 Video Study.

Howe, C., & Abedin, M. (2013). Classroom dialogue: A systematic review across four decades of research. *Cambridge Journal of Education*, *43*(3), 325–356. https://doi.org/10.1080/0305764X.2013.786024

Howe, C., Tolmie, A., Thurston, A., Topping, K., Christie, D., Livingston, K., . . . Donaldson, C. (2007). Group work in elementary science: towards organisational principles for supporting pupil learning. *Learning and Instruction*, *17*(5), 549–563.

Hox, J. J., & Kreft, I. G. G. (1994). Multilevel Analysis Methods. *Sociological Methods & Research*, *22*(3), 283–299. https://doi.org/10.1177/0049124194022003001

Huber, S. A. (2017). *Individual Student Learning Dispositions - Teacher and Student Perspectives on Incoherences and Their Development*: Technische Universität München.

Huber, S. A., Häusler, J., Jurik, V., & Seidel, T. (2015). Self-underestimating students in physics instruction: Development over a school year and its connection to internal

learning processes. *Learning and Individual Differences, 43*, 83–91. https://doi.org/10.1016/j.lindif.2015.08.021

Huber, S. A., & Seidel, T. (2018). Comparing teacher and student perspectives on the interplay of cognitive and motivational-affective student characteristics. *PloS One, 13*(8), e0200609. https://doi.org/10.1371/journal.pone.0200609

Hurrelmann, K. (2013). Das Schulsystem in Deutschland: Das "Zwei-Wege-Modell" setzt sich durch. *Zeitschrift für Pädagogik, 59*(4), 455–468.

IBM Corporation. (2016). IBM SPSS Statistics: IBM Corporation.

IBM Corporation. (2018). IBM SPSS Statistics: IBM Corporation.

Ismail, Z., & Majeed, A. (2011). Student Self Esteem and Their Perception of Teacher Behavior: A Study of Class Grouping System in Pakistan. *International Journal of Business and Social Science, 16*(2), 103–113.

Jacobs, J., Borko, H., & Koellner, K. (2009). The power of video as a tool for professional development and research: Examples from the Problem-Solving Cycle. In T. Janik & T. Seidel (Eds.), *The power of video studies in investigating teaching and learning in the classroom.* Münster: Waxmann.

Jurik, V., Gröschner, A., & Seidel, T. (2013). How student characteristics affect girls' and boys' verbal engagement in physics instruction. *Learning and Instruction, 23*, 33–42. https://doi.org/10.1016/j.learninstruc.2012.09.002

Jurik, V., Gröschner, A., & Seidel, T. (2014). Predicting students' cognitive learning activity and intrinsic learning motivation: How powerful are teacher statements, student profiles, and gender? *Learning and Individual Differences, 32*, 132–139. https://doi.org/10.1016/j.lindif.2014.01.005

Jurik, V., Häusler, J., Stubben, S., & Seidel, T. (2015). Interaction - First results of a comparative video study carried out in language and mathematics instruction. *Zeitschrift für Pädagogik, 61*(5), 692–711.

Kaiser, J., Retelsdorf, J., Südkamp, A., & Möller, J. (2013). Achievement and engagement: How student characteristics influence teacher judgments. *Learning and Instruction, 28*, 73–84. https://doi.org/10.1016/j.learninstruc.2013.06.001

Karst, K., Dotzel, S., & Dickhäuser, O. (submitted, this issue). Comparing global judgments and specific judgments of teachers about students' knowledge: Is the whole the sum of its parts? *Teaching and Teacher Education*.

Kazemi, E., & Hubbard, A. (2008). New Directions for the Design and Study of Professional Development: Attending to the Coevolution of Teachers' Participation Across Contexts. *Journal of Teacher Education*, *59*(5), 428–441. https://doi.org/10.1177/0022487108324330

Kiemer, K., Gröschner, A., Kunter, M., & Seidel, T. (2016). Instructional and motivational classroom discourse and their relationship with teacher autonomy and competence support—findings from teacher professional development. *European Journal of Psychology of Education*, 1–26. https://doi.org/10.1007/s10212-016-0324-7

Kiemer, K., Gröschner, A., Pehmer, A.-K., & Seidel, T. (2015). Effects of a classroom discourse intervention on teachers' practice and students' motivation to learn mathematics and science. *Learning and Instruction*, *35*, 94–103. https://doi.org/10.1016/j.learninstruc.2014.10.003

Kleinknecht, M., & Schneider, J. (2013). What do teachers think and feel when analyzing videos of themselves and other teachers teaching? *Teaching and Teacher Education*, *33*, 13–23. https://doi.org/10.1016/j.tate.2013.02.002

Klieme, E. (Ed.). (2010). *PISA 2009: Bilanz nach einem Jahrzehnt*. Münster: Waxmann. Retrieved from http://nbn-resolving.de/urn:nbn:de:0111-opus-35393

Klieme, E., Eichler, W., Helmke, A., Lehmann, R., Nold, G., Rolff, H.-G., . . . Willenberg, H. (Eds.). (2008). *Unterricht und Kompetenzerwerb in Deutsch und Englisch: Ergebnisse der DESI-Studie [Instruction and Competence in German and Englisch: Results of the DESI study]*. Weinheim: Beltz.

Kobarg, M., & Seidel, T. (2007). Process-Oriented Teaching – Video Analyses in High School Physics Instruction. *Unterrichtswissenschaft*, *35*(2), 148–168.

Koellner, Karen, Jacobs, Jennifer, Borko, H., Schneider, C., Pittman, Mary E., Eiteljorg, E., . . . Frykholm, J. (2007). The Problem-Solving Cycle: A Model to Support the Development of Teachers' Professional Knowledge. *Mathematical Thinking and Learning*, *9*(3), 273–303. https://doi.org/10.1080/10986060701360944

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*(4), 349–370. https://doi.org/10.1037/0096-3445.126.4.349

Kuhn, D. (2010). Teaching and learning science as argument. *Science Education*, *94*(5), 810–824. https://doi.org/10.1002/sce.20395

Kultusministerkonferenz. (2003). *Beschlüsse der Kultusministerkonferenz: Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss [Resolutions of the Conference of Ministers of Education: Educational Standards in Mathematics for the Middle School Graduation]*. Munich: Wolters Kluwer Deutschland GmbH.

Kunter, M., Dubberke, T., Baumert, Jürgen, Blum, W., Brunner, M., & Jordan, A. (2006). Mathematikunterricht in den PISA-Klassen 2004: Rahmenbedingungen, Formen und Lehr-Lernprozesse [Mathematics in the PISA 2004: general conditions, forms and teaching and learning processes]. In M. Prenzel (Ed.), *PISA 2003: Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres* (pp. 277–306). Münster, New York: Waxmann.

Lauer, P. A., Christopher, D. E., Firpo-Triplett, R., & Buchting, F. (2014). The impact of short-term professional development on participant outcomes: A review of the literature. *Professional Development in Education*, *40*(2), 207–227. https://doi.org/10.1080/19415257.2013.776619

Lee, Y., & Kinzie, M. B. (2012). Teacher question and student response with regard to cognition and language use. *Instructional Science*, *40*(6), 857–874. https://doi.org/10.1007/s11251-011-9193-2

Linnenbrink-Garcia, L., Pugh, K. J., Koskey, K. L. K., & Stewart, V. (2012). Developing Conceptual Understanding of Natural Selection: The Role of Interest, Efficacy, and Basic Prior Knowledge. *The Journal of Experimental Education*, *80*(1), 45–68. https://doi.org/10.1080/00220973.2011.559491

Lipowsky, F., Rakocy, K., Pauli, C., Reusser, K., & Klieme, E. (2007). The Same Instruction - The Same Chances for Everyone? The Distribution of Contributions from Students in Whole-Class Discussions. *Unterrichtswissenschaft*, *35*(2), 125–147.

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. https://doi.org/10.1016/j.learninstruc.2008.11.001

Lotz, Miriam, Gabriel, Katrin, Lipowsky, & Frank. (2013). Niedrig und hoch inferente Verfahren der Unterrichtsbeobachtung. Analysen zu deren gegenseitiger Validierung. *Zeitschrift für Pädagogik, 59*(3), 357–380.

Lotz, M., & Lipowsky, F. (2015). Die Hattie-Studie und ihre Bedeutung für den Unterricht: Ein Blick auf ausgewählte Aspekte der Lehrer-Schüler-Interaktion [the

Hattie study and its importance for teaching: A look at selected aspects of teacher-student interaction]. In G. Mehlhorn, K. Schöppe, & F. Schulz (Eds.), *KREAplus: Vol. 8. Begabungen entwickeln & Kreativität fördern* (pp. 97–136). München: kopaed.

Machts, N., Kaiser, J., Schmidt, F. T.C., & Möller, J. (2016). Accuracy of teachers' judgments of students' cognitive abilities: A meta-analysis. *Educational Research Review*, *19*, 85–103. https://doi.org/10.1016/j.edurev.2016.06.003

Mangold. (2014). INTERACT: Mangold International GmbH.

Marsh, H. W., Craven, R., & Debus, R. (1998). Structure, Stability, and Development of Young Children's Self-Concepts: A Multicohort-Multioccasion Study. *Child development*, *69*(4), 1030. https://doi.org/10.2307/1132361

Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *The British Journal of Educational Psychology*, *81*(Pt 1), 59–77. https://doi.org/10.1348/000709910X503501

Marsh, H. W., Trautwein, U., Ludtke, O., Koller, O., & Baumert, Jurgen. (2005). Academic self-concept, interest, grades, and standardized test scores: Reciprocal effects models of causal ordering. *Child Development*, *76*(2), 397–416. https://doi.org/10.1111/j.1467-8624.2005.00853.x

Marsh, H. W., Trautwein, U., Ludtke, O., Koller, O., & Baumert, Jurgen. (2006). Integration of multidimensional self-concept and core personality constructs: Construct validation and relations to well-being and achievement. *Journal of Personality*, *74*(2), 403–456. https://doi.org/10.1111/j.1467-6494.2005.00380.x

McDonald, M., Kazemi, Elham, & Kavanagh, S. S. (2013). Core Practices and Pedagogies of Teacher Education. *Journal of Teacher Education*, *64*(5), 378–386. https://doi.org/10.1177/0022487113493807

McElhone, D. (2012). Tell Us More: Reading Comprehension, Engagement, and Conceptual Press Discourse. *Reading Psychology*, *33*(6), 525–561. https://doi.org/10.1080/02702711.2011.561655

Mercer, N. (2008). The Seeds of Time: Why Classroom Dialogue Needs a Temporal Analysis. *Journal of the Learning Sciences*, *17*(1), 33–59. https://doi.org/10.1080/10508400701793182

Mercer, N. (2010). The analysis of classroom talk: Methods and methodologies. *The British Journal of Educational Psychology*, *80*, 1–14. https://doi.org/10.1348/000709909X479853

Mercer, N., & Dawes, L. (2014). The study of talk between teachers and students, from the 1970s until the 2010s. *Oxford Review of Education*, *40*(4), 430–445. https://doi.org/10.1080/03054985.2014.934087

Mercer, N., Dawes, L., Wegerif, R., & Sams, C. (2004). Reasoning as a scientist: ways of helping children to use language to learn science. *British Educational Research Journal*, *30*(3), 359–378.

Mercer, N., Wegerif, R., & Dawes, L. (1999). Children's Talk and the Development of Reasoning in the Classroom. *British Educational Research Journal*, *25*(1), 95–111.

Merton, R. K. (1968). The Matthew Effect in Science: The reward and communication system of science are considered. *Science*, *159*(3810), 56–63.

Meyer, H. (2016). *Was ist guter Unterricht?* (11. Auflage). Berlin: Cornelsen.

Michaelis, M. (2009). *Internes Marketing in Dienstleistungsnetzwerken: Konzeption und Erfolgsmessung [Internal marketing in service networks: conception and succes measurement]*. Zugl.: Münster (Westfalen), Univ., Diss., 2009 (1. Aufl.). *Gabler Research : Unternehmenskooperation und Netzwerkmanagement*. Wiesbaden: Gabler.

Michaels, S., & O'Connor, C. (2012). *Talk Science Primer*. Cambrigde, MA: TERC. Retrieved from http://www.cgcs.org/cms/lib/DC00001581/Centricity/Domain/155/TalkScience_Primer.pdf

Michaels, S., O'Connor, C., & Resnick, L. B. (2008). Deliberative Discourse Idealized and Realized: Accountable Talk in the Classroom and in Civic Life. *Studies in Philosophy and Education*, *27*(4), 283–297. https://doi.org/10.1007/s11217-007-9071-1

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

O'Connor, C., Michaels, S., Chapin, S., & Harbaugh, A. G. (2017). The silent and the vocal: Participation and learning in whole-class discussion. *Learning and Instruction*, *48*, 5–13. https://doi.org/10.1016/j.learninstruc.2016.11.003

OECD. (2009). *Creating Effective Teaching and Learning Environments: First Results from TALIS* (1. Aufl.). s.l.: OECD. Retrieved from http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10326180

Oliveira, A. W. (2010). Improving teacher questioning in science inquiry discussions through professional development. *Journal of Research in Science Teaching, 47*(4), 422–453. https://doi.org/10.1002/tea.20345

Osborne, Jonathan F., Henderson, J. B., MacPherson, A., Szu, E., Wild, A., & Yao, S.-Y. (2016). The development and validation of a learning progression for argumentation in science. *Journal of Research in Science Teaching*, *53*(6), 821–846. https://doi.org/10.1002/tea.21316

Pauli, C., & Lipowsky, F. (2007). Student Participation in Whole-Class Discussions and Teacher-Student Interactions in Mathematics Classrooms. *Unterrichtswissenschaft, 35*(2), 101–124.

Pehmer, A.-K., Gröschner, A., & Seidel, T. (2015a). Fostering and scaffolding student engagement in productive classroom discourse: Teachers' practice changes and reflections in light of teacher professional development. *Learning, Culture and Social Interaction, 7*, 12–27. https://doi.org/10.1016/j.lcsi.2015.05.001

Pehmer, A.-K., Gröschner, A., & Seidel, T. (2015b). How teacher professional development regarding classroom dialogue affects students' higher-order learning. *Teaching and Teacher Education, 47*, 108–119. https://doi.org/10.1016/j.tate.2014.12.007

Pielmeier, M., Böheim, R., Schindler, A.-K., Gröschner, A., Knogler, M., Alles, M., & Seidel, T. (2018). Fostering Dialogic Teaching - The "Dialogic Video Cycle" as a video based professional development programme to enhance classroom discourse. In K. Smith (Ed.), *Norsk OG Internasjonal Lærerutdanningsforskning* (pp. 63–88). Bergen: Fagbokforl.

Pielmeier, M., Huber, S. A., & Seidel, T. (2018). Is teacher judgment accuracy of students' characteristics beneficial for verbal teacher-student interactions in classroom? *Teaching and Teacher Education*. (76), 255–266. https://doi.org/10.1016/j.tate.2018.01.002

Pielmeier, M., Seidel, T., Schindler, A.-K., & Gröschner, A. (submitted). Opening 'Windows' for Teachers to Change Classroom Discourse.

Pimentel, D. S., & McNeill, K. L. (2013). Conducting Talk in Secondary Science Classrooms: Investigating Instructional Moves and Teachers' Beliefs. *Science Education*, *97*(3), 367–394. https://doi.org/10.1002/sce.21061

Praetorius, A.-K., Berner, V.-D., Zeinz, H., Scheunpflug, A., & Dresel, M. (2013). Judgment Confidence and Judgment Accuracy of Teachers in Judging Self-Concepts

of Students. *The Journal of Educational Research, 106*(1), 64–76. https://doi.org/10.1080/00220671.2012.667010

Praetorius, A.-K., Greb, K., Lipowsky, F., & Gollwitzer, M. (2010). Lehrkräfte als Diagnostiker - Welche Rolle spielt die Schülerleistung bei der Einschätzung von mathematischen Selbstkonzepten? *Journal for Educational Research Online*. (1), 121–144.

Praetorius, A.-K., Karst, K., Dickhäuser, O., & Lipowsky, F. (2011). How Teachers Rate Their Students: On Teachers' Diagnostic Competence Regarding the Academic Self-Concept. *Psychologie in Erziehung und Unterricht.* Advance online publication. https://doi.org/10.2378/peu2010.art30d

Praetorius, A.-K., Koch, T., Scheunpflug, A., Zeinz, H., & Dresel, M. (2017). Identifying determinants of teachers' judgment (in)accuracy regarding students' school-related motivations using a Bayesian cross-classified multi-level model. *Learning and Instruction.* Advance online publication. https://doi.org/10.1016/j.learninstruc.2017.06.003

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. https://doi.org/10.1016/j.learninstruc.2013.12.002

Praetorius, A.-K., & Südkamp, A. (2017). Eine Einführung in das Thema der diagnostischen Kompetenz von Lehrkräften. In A. Südkamp & A.-K. Praetorius (Eds.), *Pädagogische Psychologie und Entwicklungspsychologie. Diagnostische Kompetenz von Lehrkräften: Theoretische und methodische Weiterentwicklungen [Diagnostic Competence of Teachers: Theoretical and Methodical Evolution. Educational psychology and developmental psychology]* (pp. 13–18). Münster, New York: Waxmann.

Putnam, R. T., & Borko, H. (2000). What Do New Views of Knowledge and Thinking Have to Say About Research on Teacher Learning? *Educational Researcher, 29*(1), 4–15.

R Foundation for Statistical Computing. (2016). R Foundation for Statistical Computing.

Rakoczy, K., Klieme, E., Bürgermeister, A., & Harks, B. (2008). The Interplay Between Student Evaluation and Instruction. *Journal of Psychology, 216*(2), 111–124. https://doi.org/10.1027/0044-3409.216.2.111

Ready, D. D., & Wright, D. L. (2011). Accuracy and Inaccuracy in Teachers' Perceptions of Young Children's Cognitive Abilities: The Role of Child Background and

Classroom Context. *American Educational Research Journal*, *48*(2), 335–360. https://doi.org/10.3102/0002831210374874

Reiss, K., Sälzer, C., Schiepe-Tiska, A., Klieme, E., & Köller, O. (Eds.). (2016). *PISA 2015: Eine Studie zwischen Kontinuität und Innovation [PISA 2015: A Study between Continuity and Innovation]*. Münster, New York: Waxmann.

Resnick, L. B., Michaels, S., & O'Connor, C. (2010). How (well-structured) talk builds the mind. In D. Preiss & R. J. Sternberg (Eds.), *Innovations in educational psychology: Perspectives on learning, teaching, and human development* (pp. 163–194). New York: Springer Pub.

Reznitskaya, A., Wilkinson, I. A. G., & Oyler, J. (2016). *The Argumentation Rating Tool*.

Reznitskaya, A., Wilkinson, I. A. G., Oyler, J., Bourdage-Reninger, K., & Sykes, A. (2016, April). *Using the Argumentation Rating Tool to Support Teacher Facilitation of Inquiry Dialogue in Elementary Language Arts Classrooms.* Paper presented at the Annual Meeting of the American, Washington, DC.

Sälzer, C., & Reiss, K. (2016). PISA 2015 - die aktuelle Studie. In K. Reiss, C. Sälzer, A. Schiepe-Tiska, E. Klieme, & O. Köller (Eds.), *PISA 2015: Eine Studie zwischen Kontinuität und Innovation [PISA 2015: A Study between Continuity and Innovation]* (pp. 13–44). Münster, New York: Waxmann.

Sälzer, C., Reiss, K., Schiepe-Tiska, A., Prenzel, M., & Heinze, A. (2013). Zwischen Grundlagenwissen und Anwendungsbezug: Mathematische Kompetenz im internationalen Vergleich [Basic knowledge and application-oriented: Mathematical Competence in International Comparison]. In M. Prenzel, C. Sälzer, E. Klieme, & O. Köller (Eds.), *Pisa 2012: Fortschritte und Herausforderungen in Deutschland* (pp. 47–97). Münster: Waxmann.

Santagata, R. (2009). Designing Video-Based Professional Development for Mathematics Teachers in Low-Performing Schools. *Journal of Teacher Education*, *60*(1), 38–51. https://doi.org/10.1177/0022487108328485

Schiepe-Tiska, A., & Schmidtner, S. (2013). Mathematikbezogene emotionale und motivationale Orientierungen, Einstellungen und Verhaltensweisen von Jugendlichen in PISA 2012 [Mathematics-related Emotional and Motivational Orientations, attitudes and Behaviors of Adolescents in PISA 2012]. In M. Prenzel, C. Sälzer, E. Klieme, & O. Köller (Eds.), *Pisa 2012: Fortschritte und Herausforderungen in Deutschland* (pp. 99–121). Münster: Waxmann.

Schindler, A.-K., Böheim, R., Pielmeier, M., Gröschner, A., & Seidel, T. (2018). *Videoinstrument "Dialogisches Lehren und Lernen": Zur Auswertung der DIALOGUE II Unterrichtsvideos*. München.

Schindler, A.-K., Gröschner, A., & Seidel, T. (2016). Teaching Science Effectively: A Case Study on Student Verbal Engagement in Classroom Dialogue. *ORBIS SCHOLAE, 9*(2), 9–34. https://doi.org/10.14712/23363177.2015.78

Schindler, A.-K., Seidel, T., Böheim, R., Pielmeier, M., Knogler, M., Alles, M., & Gröschner, A. (submitted). The Practice Challenge Classroom Discourse: Understanding Teachers' Individual Learning Zones and Trajectories of Change during Teacher Professional Development.

Schrader, F.-W. (2013). Teacher Diagnosis and Diagnostic Competence. *Beiträge zur Lehrerinnen- und Lehrerbildung*, *31*(2), 154–165.

Schrader, F.-W. (2017). Diagnostische Kompetenz von Lehrkräften - Anmerkungen zur Weiterentwicklung des Konstrukts. In A. Südkamp & A.-K. Praetorius (Eds.), *Pädagogische Psychologie und Entwicklungspsychologie. Diagnostische Kompetenz von Lehrkräften: Theoretische und methodische Weiterentwicklungen [Diagnostic Competence of Teachers: Theoretical and Methodical Evolution. Educational psychology and developmental psychology]* (pp. 247–255). Münster, New York: Waxmann.

Scott, Philip, Mortimer, E. F., & Aguiar, O. G. (2006). The tension between authoritative and dialogic discourse: A fundamental characteristic of meaning making interactions in high school science lessons. *Science Education*, *90*(4), 605–631. https://doi.org/10.1002/sce.20131

Sedova, K., & Salamounova, Z. (2016). Teacher expectancies, teacher behaviour and students' participation in classroom discourse. *Journal of Educational Enquiry*, *15*(1), 44–61. Retrieved from http://www.ojs.unisa.edu.au/index.php/EDEQ/article/viewFile/1304/898

Sedova, K., Sedlacek, M., & Svaricek, R. (2016). Teacher professional development as a means of transforming student classroom talk. *Teaching and Teacher Education*, *57*, 14–25. https://doi.org/10.1016/j.tate.2016.03.005

Seidel, T. (2006). The role of student characteristics in studying micro teaching–learning environments. *Learning Environments Research*, *9*(3), 253–271. https://doi.org/10.1007/s10984-006-9012-x

Seidel, T. (2014). Lehrerhandeln im Unterricht [Teacher Behavior in classroom]. In E. Terhart, H. Bennewitz, & M. Rothland (Eds.), *Handbuch der Forschung zum Lehrerberuf* (2nd ed., pp. 781–806). Münster, New York: Waxmann.

Seidel, T., & Prenzel, M. (2006). Stability of teaching patterns in physics instruction: Findings from a video study. *Learning and Instruction*, *16*(3), 228–240. https://doi.org/10.1016/j.learninstruc.2006.03.002

Seidel, T., Prenzel, M., & Kobarg, M. (Eds.). (2005). *How to run a video study: Technical report of the IPN Video Study*. Münster: Waxmann.

Seidel, T., Rimmele, Rolf, Prenzel, & Manfred. (2003). Gelegenheitsstrukturen beim Klassengespräch und ihre Bedeutung für die Lernmotivation. Videoanalysen in Kombination mit Schülerselbsteinschätzungen. *Unterrichtswissenschaft*, *31*(2), 142–165.

Seidel, T., & Shavelson, R. (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. https://doi.org/10.3102/0034654307310317

Seidel, T., Stürmer, K., Blomberg, G., Kobarg, M., & Schwindt, K. (2011). Teacher learning from analysis of videotaped classroom situations: Does it make a difference whether teachers observe their own teaching or that of others? *Teaching and Teacher Education*, *27*(2), 259–267. https://doi.org/10.1016/j.tate.2010.08.009

Shavelson, R., & Stern, P. (1981). Research on Teachers' Pedagogical Thoughts, Judgments, Decisions, and Behavior. *Review of Educational Research*, *51*(4), 455–498. https://doi.org/10.3102/00346543051004455

Sherin, Miriam, & van Es, Elizabeth. (2002). Using Video to Support Teachers' Ability to Interpret Classroom Interactions. In D. A. Willis, J. Price, & N. Davis (Eds.), *Society for Information Technology & Teacher Education International Conference 2002* (pp. 2532–2536). Nashville, Tennessee, USA: Association for the Advancement of Computing in Education (AACE). Retrieved from https://www.learntechlib.org/p/17741

Shulman, L. (1987). Knowledge and Teaching: Foundations of the New Reform. *Harvard Educational Review*, *57*(1), 1–23. https://doi.org/10.17763/haer.57.1.j463w79r56455411

Snow, R. E., Corno, L., & Jackson, D., III. (1996). Individual Differences in Affective and Conative Functions. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of Educational Psychology* (pp. 243–310). New York: Macmillan.

Spinath, B. (2005). Accuracy of Teacher Judgments on Student Characteristics and the Construct of Diagnostic Competence. *German Journal of Educational Psychology*, *19*(1/2), 85–95.

Staatsministerium für Unterricht und Kultus. (2004). 8 Mathematik. Retrieved from http://www.isb-gym8-lehrplan.de/contentserv/3.1.neu/g8.de/index.php?StoryID=26279

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, *XXI*(4), 360–407.

Stürmer, K., & Seidel, T. (2015). Assessing Professional Vision in Teacher Candidates. *Zeitschrift f?r Psychologie*, *223*(1), 54–63. https://doi.org/10.1027/2151-2604/a000200

Stürmer, K., Seidel, T., Müller, K., Häusler, J., & S. Cortina, K. (2017). What is in the eye of preservice teachers while instructing?: An eye-tracking study about attention processes in different teaching situations. *Zeitschrift für Erziehungswissenschaft*, *20*(S1), 75–92. https://doi.org/10.1007/s11618-017-0731-9

Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, *104*(3), 743–762. https://doi.org/10.1037/a0027627

Südkamp, A., & Praetorius, A.-K. (Eds.). (2017). *Pädagogische Psychologie und Entwicklungspsychologie. Diagnostische Kompetenz von Lehrkräften: Theoretische und methodische Weiterentwicklungen [Diagnostic Competence of Teachers: Theoretical and Methodical Evolution. Educational psychology and developmental psychology]*. Münster, New York: Waxmann.

Südkamp, A., Praetorius, A.-K., & Spinath, B. (2017). Teachers' judgment accuracy concerning consistent and inconsistent student profiles. *Teaching and Teacher Education.* Advance online publication. https://doi.org/10.1016/j.tate.2017.09.016

Thiede, K. W., Brendefur, J. L., Carney, M. B., Champion, J., Turner, L., Stewart, R., & Osguthorpe, R. D. (2018). Improving the accuracy of teachers' judgments of student learning. *Teaching and Teacher Education*, *76*, 106–115. https://doi.org/10.1016/j.tate.2018.08.004

Timperley, Helen. (2013). Feedback. In J. A. C. Hattie & E. M. Anderman (Eds.), *Educational Psychology Handbook. International Guide to Student Achievement* (pp. 402–404). Hoboken: Taylor and Francis.

Turner, J., Meyer, D., Midgley, C., & Patrick, H. (2003). Teacher discourse and sixth graders' reported affect and achievement behaviors in two high-mastery/high-performance mathematics classrooms. *The Elementary School Journal*, *103*(4), 359–382.

Turner, J., Warzon, K. B., & Christensen, A. (2011). Motivating Mathematics Learning: Changes in Teachers' Practices and Beliefs During a Nine-Month Collaboration. *American Educational Research Journal*, *48*(3), 718–762. https://doi.org/10.3102/0002831210385103

Van de Pol, J., Brindley, S., & Higham, Rupert J. E. (2017). Two secondary teachers?: Understanding and classroom practice of dialogic teaching: a case study. *Educational Studies*, *18*(1), 1–19. https://doi.org/10.1080/03055698.2017.1293508

Van den Bergh, L., Ros, A., & Beijaard, D. (2015). Teacher learning in the context of a continuing professional development programme: A case study. *Teaching and Teacher Education*, *47*, 142–150. https://doi.org/10.1016/j.tate.2015.01.002

Van den Bogert, N., van Bruggen, J., Kostons, D., & Jochems, W. (2014). First steps into understanding teachers' visual perception of classroom events. *Teaching and Teacher Education*, *37*, 208–216. https://doi.org/10.1016/j.tate.2013.09.001

Van der Veen, C., Dobber, M., & van Oers, B. (2017). Engaging Children in Dialogic Classroom Talk: Does It Contribute to a Dialogical Self? In F. Meijers & H. Hermans (Eds.), *The Dialogical Self Theory in Education* (pp. 49–63). Cham: Springer International Publishing.

Van Es, Elizabeth, & Sherin, Miriam G. (2006). How Different Video Club Designs Support Teachers in "Learning to Notice". *Journal of Computing in Teacher Education*, *22*(4), 125–135.

Van Es, Elizabeth A., Tunney, J., Goldsmith, L. T., & Seago, N. (2014). A Framework for the Facilitation of Teachers' Analysis of Video. *Journal of Teacher Education*, *65*(4), 340–356. https://doi.org/10.1177/0022487114534266

Van Zee, E. H., Iwasky, M., Kurose, A., Simpson, D., & Wild, J. (2001). Student and Teacher Questioning during Conversations about Science. *Journal of Research in Science Teaching*, *38*(2), 159–190.

Vermunt, J. D., & Endedijk, M. D. (2011). Patterns in teacher learning in different phases of the professional career. *Learning and Individual Differences*, *21*(3), 294–302. https://doi.org/10.1016/j.lindif.2010.11.019

Vigerske, S. (2017). *Transfer von Lehrerfortbildungsinhalten in die Praxis: Eine empirische Untersuchung zur Transferqualität und zu Einflussfaktoren*. Retrieved from http://dx.doi.org/10.1007/978-3-658-17685-3

Voerman, L., Meijer, P. C., Korthagen, F. A.J., & Simons, R. J. (2012). Types and frequencies of feedback interventions in classroom interaction in secondary education. *Teaching and Teacher Education*, *28*(8), 1107–1115. https://doi.org/10.1016/j.tate.2012.06.006

Voss, T., Kunina-Habenicht, O., Hoehne, V., & Kunter, M. (2015). Teachers' pedagogical knowledge: Empirical approaches and findings. *Zeitschrift für Erziehungswissenschaft*, *18*(2), 187–223. https://doi.org/10.1007/s11618-015-0626-6

Vrikki, M., Warwick, P., Vermunt, J. D., Mercer, N., & van Halem, N. (2017). Teacher learning in the context of Lesson Study: A video-based analysis of teacher discussions. *Teaching and Teacher Education*, *61*, 211–224. https://doi.org/10.1016/j.tate.2016.10.014

Walshaw, M., & Anthony, G. (2008). The Teacher's Role in Classroom Discourse: A Review of Recent Research Into Mathematics Classrooms. *Review of Educational Research*, *78*(3), 516–551. https://doi.org/10.3102/0034654308320292

Wang, X. (2013). Why Students Choose STEM Majors. *American Educational Research Journal*, *50*(5), 1081–1121. https://doi.org/10.3102/0002831213488622

Webb, N. M. (2009). The teacher's role in promoting collaborative dialogue in the classroom. *The British Journal of Educational Psychology*, *79*(Pt 1), 1–28. https://doi.org/10.1348/000709908X380772

Webb, N. M., Franke, M. L., Ing, M., Wong, J., Fernandez, C. H., Shin, N., & Turrou, A. C. (2014). Engaging with others' mathematical ideas: Interrelationships among student participation, teachers' instructional practices, and learning. *International Journal of Educational Research*, *63*, 79–93. https://doi.org/10.1016/j.ijer.2013.02.001

Wikipedia. (2019). Teaching method - Wikipedia, the free encyclopedia. Retrieved from https://en.wikipedia.org/wiki/Teaching_method (accessed 09-June-2019)

Wilkinson, I. A. G., Reznitskaya, A., Bourdage, K., Oyler, J., Glina, M., Drewry, R., . . . Nelson, K. (2016). Toward a more dialogic pedagogy: Changing teachers? beliefs and

practices through professional development in language arts classrooms. *Language and Education*, *31*(1), 65–82. https://doi.org/10.1080/09500782.2016.1230129

Yamamoto, T., & Imai-Matsumura, K. (2015). Teachers' Gaze and Awareness of Students' Behavior: Using An Eye Tracker. *Comprehensive Psychology*, *2*(6), 1-7. https://doi.org/10.2466/01.IT.2.6

Yin, R. K. (2014). *Case study research: Design and methods* (5. edition). Los Angeles, London, New Delhi, Singapore, Washington, DC: SAGE.

Zhang, M., Lundeberg, M., Koehler, M. J., & Eberhardt, J. (2011). Understanding affordances and challenges of three types of video for teacher professional development. *Teaching and Teacher Education*, *27*(2), 454–462. https://doi.org/10.1016/j.tate.2010.09.015

List of Abbreviations

DFG	Deutsche Forschungsgesellschaft [German Research Fundation]
DVC	Dialogic Video Cycle
ICC	intra-class correlation
IRF	Initiation – response – feedback
NeDiKo	Netzwerk – Diagnostische Kompetenz [Network – Diagnostic Competences]
OECD	Organization for Economic Co-operation and Development
PISA	Programme for International Student Assessment
PRE	Pre-achievement
QD	Quality dimension
SC	Self-concept of ability
TJA	Teacher judgment accuracy
TPD	Teacher professional development

List of Figures

List of Tables

Appendix

1 Supplement 1: Statement of Distribution of Tasks and Responsibilities

Supplement 1: Statement of Distribution of Tasks and Responsibilities

2 Supplement 2: Study I (Published Manuscript)

Pielmeier, M., Huber, S., & Seidel, T. (2018). Is teacher judgment accuracy of students' characteristics beneficial for verbal teacher-student interactions in classroom? *Teaching and Teacher Education*. (76), 255–266. https://doi.org/10.1016/j.tate.2018.01.002.

Supplement 2: Study I (Published Manuscript)

3 Supplement 3: Study II (Manuscript submitted for publication)

Pielmeier, M., Seidel, S., Schindler, A.-K., & Gröschner, A. (submitted). Opening 'Windows' for Teachers to Change Classroom Discourse. *Teaching and Teacher Education.*

4 Supplement 4: Eidesstattliche Erklärung