Instructional Strategies in Teacher Education

Comparing two Video-Based Courses on Classroom Management

Tom Schep

Inauguraldissertation

zur Erlangung des Grades eines Doktors der Philosophie (Dr. Phil.) im Fachbereich Erziehungswissenschaften der Johann-Wolfgang-Goethe-Universität zu Frankfurt am Main

Erster Gutachter: Prof. Dr. Udo Rauin, Goethe-Universität Frankfurt am Main
Zweite Gutachterin: Prof.'in Dr. Ilonca Hardy, Goethe-Universität Frankfurt am Main

Tag der mündlichen Prüfung: 11. Dezember 2018

D30

This work is dedicated to Olivia, David, Vera, and Christiane

ABSTRACT

Universities typically offer video-based courses and internships to prepare prospective teachers for classroom management. Empirical studies have shown the effectiveness of analyzing video clips in programs for teacher training. However, no empirical studies have explored the relationship between the teaching experience of prospective teachers and the effectivity of video-based courses on classroom management. The research question was: *How can classroom management competency be effectively conveyed in video-based courses regarding the differences in practical experiences of prospective teachers.* Theory on cognitive load and situated learning suggests that prospective teachers with teaching experience could benefit more from a situated instructional strategy, whereas prospective teachers without this experience benefit more from a cognitive instructional strategy.

The sample consisted of a total of 87 prospective teachers belonging to a first cohort that had participated in an internship and a second cohort that had not yet participated. Participants of both cohorts were randomly assigned to a course following one of the two instructional strategies. A video test was designed to assess classroom management competence. Additionally, measurements of prospective teachers' attentiveness and their self-efficacy at the start of the internship were included.

The study showed a significant main effect between pretest and posttest. However, neither a significant difference in the effectiveness between the two instructional strategies nor between the two cohorts was found (first order). Also, the effectiveness of the instructional strategy was not influenced by the availability of teaching experience (second order). Scores on the video test were not significantly related to prospective teachers' attentiveness. The prospective teachers that participated in the course before participation in the internship rated their own classroom management significantly lower than the prospective teachers that participated in the course after participating in their internship.

These results firstly indicate that there was no significant difference in cognitive load between the groups, implying that both instructional strategies are suitable for the conveying of classroom management competence within universities. Secondly, the current study has found no evidence that the trait of attentiveness influences prospective teachers' performance within video tests. Thirdly, these results give reason to believe that the valid assessment of self-efficacy on classroom management depends on the degree of familiarity with the topic, for example, through previous dealings within video-based courses. The limited reliability of the new video test should be taken into account when interpreting the results of the current study.

ACKNOWLEDGMENTS

I wish to express my deepest gratitude to my major supervisor, Prof. Dr. Udo Rauin, for all of his support, guidance, feedback, and the stimulating discussions throughout this project. I am also very grateful to Prof.'in Dr. Ilonca Hardy for her support as my second supervisor.

I would like to thank my (former) colleagues, Dr. Jette Horstmeyer, Dr. Matthias Herrle, Dr. Johannes Appel, Dr. Sebastian Stehle, Benjamin Tolweth, Andrew Seifert, Sebastian Breitenbach, and Dr. Sebastian Poloczek. You have given me invaluable feedback throughout the different phases of my project and on the manuscript.

Furthermore, I would like to thank my student assistants, Julian Marx, Sherouk El Sheshai, Rebecca Vainstain, Ivo Greuloch, and Belkacem Jellouli, for their assistance in coding and support in the project. I would especially like to thank Louisa Woller for her committed coding and valuable feedback on the manuscript.

Special thanks go out to my brother Tiko Schep and my dear friends Inge van de Wiel, Martijn Meerhoff, Jasper Beckeringh, and Nathan Ducastel for their advice, feedback on the manuscript and support.

Finally, I would like to express my deepest respect to my wife, Christiane. Her understanding, patience, confidence, flexibility, and pragmatism during a strenuous period for our family have helped me in countless ways and were invaluable for completing this project.

Thank you.

LIST OF TABLES

LIST OF FIGURES

TABLE OF CONTENTS

1. Introduction

1.1 Programs for Teacher Education: Value Added and Reality Shock

Recent empirical pedagogical research has shown the relative importance of teaching quality for students' learning outcomes (Hattie, 2009; Lipowsky, 2006; Seidel & Shavelson, 2007). With this discovery, the assessment of the effectiveness or value-added of programs and interventions for teacher education moved into the central focus of both policy (KMK, 2004) and pedagogical research (Cochran-Smith & Zeichner, 2005; Rauin Udo, 2011; Terhart, Bennewitz, & Rothland, 2011). However, the effectiveness of teacher education is not self-evident (Darling-Hammond, 1999; Kennedy, Soyeon, & Jonyoung, 2008).

The transition between university-based teacher education and the teaching practice is often characterized as a reality shock (Müller-Fohrbrodt & Dann, 1978; Veenman, 1984). This means that prospective teachers do not feel that the programs sufficiently prepare them for the teaching practice and that the acquired knowledge remains largely useless (Anselm, 2010). Prospective teachers have shown to be able to reproduce curricular knowledge in tests. However, they have difficulties in transferring this knowledge to the teaching practice. This knowledge then becomes inert knowledge (Renkl, 1996). Instead of applying the knowledge learned within programs for teacher training, beginning teachers tend to teach like they were taught themselves at school (Hascher, 2011).

It remains a continuous debate among teacher trainers and pedagogical researchers, to what degree and under which conditions starting teachers can successfully apply theoretical knowledge gained in universities to their future teaching practice (Korthagen & Kessels, 1999; Mägdefrau & Schumacher, 2001). These results pose questions regarding the learning mechanisms of prospective teachers and how to initiate relevant connections between university-based courses and practical experiences (Zeichner, 2009).

1.2 Classroom Management

The challenge of finding effective methods for preparing prospective teachers is especially relevant to the aspect of classroom management (Jones, 2006 Havers, 2010). The discussions on

the topic of issues on students' discipline and teachers' leadership go back to the introduction of compulsory education in the nineteenth century (Haag & Streber, 2012). However, empirical research on classroom management has its origin in educational research in the USA in the second half of the 20th century (Brophy, 2006; Evertson & Weinstein, 2006b). In the 'Handbook on Classroom Management,' Evertson & Weinstein define classroom management as "the actions teachers take to create an environment that supports and facilitates both academic and social-emotional learning" (Evertson & Weinstein, 2006b, p. 4). From this definition, it follows firstly, that classroom management focuses on a teacher's in-situ behavior, rather than the teacher's ability to plan lessons or design learning materials. Secondly, this definition shows that classroom management consists of the creation of the conditions for effective learning, rather than the learning activities themselves. Therefore, it is assumed that classroom management represents a teacher ability that is independent of the teaching subject. Thirdly, it follows from this definition that classroom management is considered supportive of learning, but it does not represent a learning objective in itself, as opposed to the teaching of discipline.

Several meta-studies have shown the importance of teachers' ability to manage classrooms (Seidel & Shavelson, 2007; Wang, Haertel, & Walberg, 1993). When a teacher can effectively strengthen classroom order and prevent lesson disturbances, he or she can increase the available learning time (Helmke, 2007) and student engagement (Emmer & Stough, 2001). Consequently, effective classroom management has gained a central influence within models of teaching quality (Helmke, 2012; Ophardt & Thiel, 2008). Also, teachers themselves have an interest in being able to deal with issues regarding classroom order. Empirical research has shown that higher degrees of lesson disturbances are associated with low levels of teachers' self-efficacy and well-being and with high levels of teacher's burn-out and early drop-out (Brouwers & Tomic, 2000; Friedman, 2006; König & Rothland, 2016).

Problems related to classroom order seem especially acute for starting teachers (Veenman, 1984). Many teachers feel that they are inadequately prepared to deal with student disturbances (Doyle, 2006) and that a more thorough preparation in programs for teacher education would reduce the stress related to the first period of their professional careers (Merrett & Wheldall, 1993). This implies that if programs of teacher education can find ways to convey professional

knowledge on classroom management effectively, the reality shock can be significantly reduced.

1.3 Classroom Management in Programs of Teacher Education

The topic of classroom management has found its way into most university-based programs for teacher education (Lohmann, Seidel, & Terhart, 2011), and most teacher educators consider the topic important for prospective teachers (Kunina-Habenicht et al., 2012). Typical initiatives for bridging the gap between theory and practice focus on the strengthening of internships and the analysis of video recordings.

The expectations of the integration of teaching experiences in programs of teacher education have always been very high (Hascher, 2012b; Holtz, 2014). However, empirical research on the effects of internships is fragmentary, mostly based on self-reports, and shows a multitude of effects, both desired and undesired (Hascher, 2012a). Moreover, because the degree of autonomous teaching performed by prospective teachers strongly varies, internships cannot be conceived as a uniform and consistent part of programs of teacher education (Gröschner et al., 2015). For the effective conveying of knowledge on classroom management, an important role is ascribed to the reflection on personal teaching experiences (Larrivee, 2006; Ophardt & Thiel, 2013). However, to the knowledge of the author of the current study, the specific effects of internships on prospective teachers' classroom management competence have not been systematically investigated.

Another typical method of connecting theory to practice in university-based programs for teacher education is the analysis of video-recorded cases (Blomberg, Renkl, Sherin, Borko, & Seidel, 2013; Jones, 2006; Krammer, 2014; Santagata & Guarino, 2011). This method assumes that video recordings provide a 'window' into authentic teaching (Krammer & Reusser, 2004) that allows for the theory-based reflection and analysis, without the pressure of immediate agency (Helmke & Helmke, 2004). In recent years, many studies have shown the effectiveness of video-based case analysis on a multitude of competence-aspects like knowledge, ability, beliefs, and reflection (Gaudin & Chaliès, 2015; Seidel & Thiel, 2017; Steffensky & Kleinknecht, 2016). Several studies have also shown positive effects on prospective teachers' classroom management competence (Gold, Hellermann, & Holodynski, 2017; Kramer, König, Kaiser, Ligtvoet, & Blömeke,

2017).

1.4 Video-based Case Analysis: Cognitive or Situated Instructional Strategies

As these studies have shown the potential for video-based case analysis, recent empirical studies focused on the specific question of how videos should be didactically embedded within the courses (Blomberg et al., 2013). The core of this question focuses on the relation between the case portrayed in the video clip and the general rule or theoretical principle that that video clip is supposed to represent. More specifically, this research aims to compare if video clips should provide an anchor that activates acquired experiences, beliefs, and tacit knowledge (Neuweg, 2004) from which regularities and theories should subsequently be deduced. Or that videos should alternatively be utilized as a means to apply previously conveyed theoretical concepts through a fixed set of strongly structured steps (Seidel, Blomberg, & Renkl, 2013).

This question is closely related to the debate on the degree of cognitive load associated with the analyzing of video clips. Research shows that novice teachers tend to have difficulties in dealing with the complexity of video recordings (Carter, Cushing, Sabers, Stein, & Berliner, 1988; Sabers, Cushing, & Berliner, 1991). It is argued that learning environments using videos should, therefore, follow a *cognitive* instructional strategy and be video-cases should be analyzed by closely following pre-defined steps to reduce cognitive load (Sweller, 1988). An alternative approach sees videos as a way to immerse learners within the complexity of the classroom setting. This initial complexity should not be reduced as this would distance the social learning context from the social context in which the knowledge should be applied, increasing the risk of flawed knowledge transfer (Greeno, 1998). Therefore, courses should follow a *situated* instructional strategy in which learning is situated as closely as possible to its original teaching context and learners should be encouraged to use videos to activate and reflect their own experience (Korthagen, 2010).

To date, the empirical research that systematically compares the effects of both instructional designs is limited (Blomberg, Sherin, Renkl, Glogger, & Seidel, 2014; Kumschick et al., 2017; Seidel et al., 2013; Syring et al., 2015). First studies indicate that both instructional strategies can be effective, but their relative impact depends on the specific competence types: knowledge tends

to be more strongly increased through a cognitive instructional strategy, whereas context-related competence types, like for example the ability to plan lessons, benefit more from a situated instructional strategy (Seidel et al., 2013). For the specific teaching aspect of classroom management, the research is fragmentary (Kumschick et al., 2017; Syring et al., 2015).

1.5 Classroom Management Competence

To theoretically discuss these different learning effects on prospective teachers, a fitting model of teachers' competence is needed. The concept of competence is generally understood as "complex ability constructs that are context-specific, trainable, and closely related to real life" (Koeppen, Hartig, Klieme, & Leutner, 2008, p. 61). The concept of competence is typically used for the measurement of teachers' professional ability and for the identification of effective opportunities to learn (Klieme, Hartig, & Rauch, 2008).

Recent initiatives aim at developing models that are more sensitive to the context of teaching. The PID-model, for example, defines competence as a continuum where mental *dispositions* and real-life *performance* represent the two ends. So-called *situation-specific skills* form the middle ground, bridging the gap between the two ends of the continuum (Blömeke, Gustafsson, & Shavelson, 2015b). These situation-specific skills are assumed to represent teachers' skills that are on the one hand based on their mental disposition like knowledge, beliefs, and motivation and are on the other hand predictive for successful real-life performance. Situation-specific skills are activated through simulations of professional practice, for example in video clips (Stahnke, Schueler, & Roesken-Winter, 2016). The PID-model is used to theorize the effects of courses in programs for prospective teachers, where the teaching practice is often only available through their simulations in video clips.

Situated models of competence are typically measured with instruments using video clips as item-prompts (Borko, 2016; Kersting, 2008; Lindmeier, 2013; Seidel & Thiel, 2017). In these video tests, short video clips are used to activate test-participants' situation-specific skills. Rating-items (Seidel, Blomberg, & Stürmer, 2010) or open-response items (Blömeke, König, Suhl, Hoth, & Döhrmann, 2015) are used to quantify prospective teachers' situation-specific skills. Also for the measurement of classroom management competence, several video tests were developed

(Barth, 2016; Gold, Hellermann, & Holodynski, 2016; König, 2015b).

1.6 Aims of the Current Study

The current study aims to investigate how video-based courses can build on the practical experiences of prospective teachers when conveying classroom management competence. The study consists of a quasi-experimental research design in which the learning effects of four video-based courses on classroom management are measured in pretest and posttest. For these four video-based courses, two factors were systematically controlled. One cohort of prospective teachers participated in the video-based courses before starting with an internship. Another cohort participated in the video-based courses after completion of the internship. Each cohort consisted of two groups who were randomly assigned to either a course, using a situated instructional approach, or to a course using a cognitive instructional approach. Classroom management competence was measured with a newly developed video test.

The current study aims to make contributions on three levels. On a practical level, the study aims to contribute to the discussion on how to effectively convey classroom management competence of prospective teachers through video-based courses. On a theoretical level, the study aims to further the discussion on how prospective teachers' classroom management competence should be modeled. On a methodological level, the study aims to contribute to the way prospective teachers' situated competence should be validly measured.

1.7 Structure of the Manuscript

Regarding these aims, the manuscript reporting of the current study starts with a theoretical exploration of the teaching aspect of classroom management (2). It is discussed why teachers are confronted with the challenge of dealing with keeping classroom order in the first place and why the classroom social setting typically poses challenges for teachers (2.1). Educational research on classroom management is discussed to provide a better understanding of what classroom management refers to and what teachers' effective classroom management strategies look like (2.2). This research is typically grouped in three approaches: the behavioristic, ecological and process-outcome approach. These approaches, their respective theoretical assumptions, its

supporting empirical findings and their relevance for the different classroom management strategies are then discussed in more detail. The chapter is concluded with a definition of the understanding of classroom management (2.3).

After defining the concept of classroom management, chapter 3 discusses the way classroom management is typically conceptualized in models of teacher competence. The chapter starts by defining a general understanding of the meaning, its function and general assumption of the concept of competence (3.1). Then, several sections discuss three alternative understandings of teacher competence and their respective understandings of classroom management competence: professional competence (3.2), professional vision (3.3) and the PID-model (3.4). For illustration and further reference, two examples of empirical studies that have operationalized and assessed classroom management competence (3.5) are discussed. Then, the role of the capacity to process information on performance on video tests is discussed (3.6), after which the importance of subjective ratings of self-efficacy are laid out (3.7). The chapter is concluded with a section summarizing the main points of the chapter and stating their relevance for the current study (3.8).

Chapter four discusses several ways through which prospective teacher typically acquire classroom management competence in university-based courses. The chapter starts by discussing general guidelines for the design of courses on classroom management in universities (4.1). The chapter continues with a discussion of the effects of internships on several competence aspects and how they are relevant for the acquiring of classroom management competence (4.2). The method of the analysis of video-cases is typically ascribed strong potential for programs of teacher education. Section 4.3 describes the specifics of this method and the guidelines for the design of video-based courses for prospective teachers and includes the discussion on the details regarding a situated and cognitive instructional strategy: their theoretical assumptions and their relevance for learning video-based learning. It is explained how both instructional strategies relate to the challenges associated with learning from video-cases. Additionally, four studies are compared that systematically evaluate the effects of both instructional strategies in video-based courses in teacher education. In the concluding section of this chapter (4.4), it is explained why the use of situated and cognitive instructional strategies is expected to make effective use of the

differences between prospective teachers resulting from participation in an internship.

In chapter five, the most important points of the preceding chapters are summarized and the respective research desiderata are listed (5.1, 5.2, and 5.3) Building on these desiderata, the research questions are formulated (5.4).

In chapter 6, the methodology with which the research question is answered is laid out. Chapter six starts by giving an overview of the relevant theory on test-development (6.1). Then the process of the development of a new video test is described in more detail (6.2) by describing the process of the selection of video clips and the Item-development. Here, also the results of a pilot-assessment and the changes that were made to the test based on the result of pilot-assessment are presented. Additionally, the development of the coding rubric and the coding process is described. Chapter 6.3 describes the development process and content of the treatments. Apart from the newly developed video test, several additional measuring instruments were used, for example, for the assessment of construct validity. Chapter 6.4 lists these additional instruments and discusses their functions within the current study. The subsequent sections describe the sample (6.5), the research ethics (6.6) the statistical hypotheses that were tested (6.7), and the applied methods for data-analysis (6.8).

The results of the current study are described in chapter 7. The chapter starts by reporting the results of inter-rater reliability (7.1). The subsequent section describes some final revisions that were made to the video test based on the result of item-analysis (7.2). In chapter 7.3, the factorial structure of the video test is explored. Chapter 7.4 presents the descriptive summaries of all data of the different instruments. Chapter 7.5 discusses the results of the assessment of reliability and construct validity. Chapter 7.6 describes the results of course-evaluations of prospective teachers. In the final section of the chapter, the results of the data analysis on the treatment effects as laid out in the research questions are presented (7.7).

In the final chapter, the results presented in chapter 7 are discussed (8). The chapter starts by reviewing the results regarding the research question (8.1). In section 8.2, the limitations regarding the instrument, the sample, and the treatment are discussed. In the final section of this chapter (8.3), the relevance of the findings on the and the possibilities for further research

are evaluated regarding the main aims of the current study: teaching on classroom management competence in university, the modeling of situated aspects of classroom management competence and its measurements with video tests.

2. Classroom Management

Textbooks on classroom management focus on a multitude of teaching aspects and intervention strategies. Topics include a focus on different disciplinary measures (Canter, 2010), the improvement of teacher-student relationship (Mayr, 2006), the establishment of a productive learning environment (Good & Brophy, 2008), the physical arrangement of the classroom (Borich, 2011), the effective use of available learning time (Seidel, 2009), and the effective use of rules and procedures (Ophardt & Thiel, 2013). These textbooks can serve as a good resource for understanding how teachers can increase their skill in keeping classroom order. However, the variety of presented strategies is very broad, implying that classroom order refers to a very broad and heterogeneous teaching aspect. To develop a more thorough understanding of how all these strategies can improve classroom order, we must take a closer look at the educational research on the topic of classroom management. This chapter aims at providing a systematic overview of the educational research on classroom management by providing answers to the following questions. Why is managing classroom order so challenging? What are the main academic perspectives on effective classroom management? Moreover, how do these perspectives help to understand what classroom management looks like in a concrete teaching situation?

This chapter answers these questions by firstly discussing some of the general characteristics of the classroom and including why these characteristics create the teacher's challenge of keeping classroom order (2.1). Secondly, the educational research on the topic of classroom management is summarized by discussing the three main academic approaches to the topic of classroom management: the *behavioristic*, the *ecological* and the *process-outcome* approach (2.2). Finally, the main findings of this chapter are summarized, and their relevance is discussed for the prospective teachers' acquiring of classroom management competence (2.3).

2.1 Teachers' Challenge to Manage Classroom Order

In the introduction we have presented a definition of classroom management, stating that classroom management refers to "the actions teachers take to create an environment that supports and facilitates both academic and social-emotional learning." (Evertson & Weinstein,

2006a, p. 4). This definition implies that classroom management focuses on teacher behavior instead of student behavior and that it aims at creating a supportive learning environment for the whole class rather than dealing with disturbances of individual students. However, this definition does not help to understand why managing classrooms can be so challenging. To answer this question, the concept of classroom order is placed into theoretical context by relating it to the concept of student discipline (2.1.1), by describing some of the core qualities of classroom settings from an interaction perspective (2.1.2), and by discussing students' and teachers' perspective on classroom management (2.1.3). In the last paragraph (2.1.4) the main points are summarized.

2.1.1 Students' misbehavior.

The challenge of dealing with and controlling of children's misbehavior has been a theme for pedagogical thought throughout the centuries (Keller, 2008; Tenorth, 2010). The challenge of dealing with issues related to students' discipline becomes the responsibility of the state with the introduction of compulsory education at the beginning of the 19th century (Haag & Streber, 2012). Up to then, children's upbringing and socialization were mainly seen as the responsibility of the (extended) family (Haag & Streber, 2012, p. 36). With the introduction of compulsory education, this responsibility shifted from the family to the state. From this moment on, schools and teachers were appointed with the task to deal with relatively large and heterogeneous groups of children and to socialize them in such a way that state-defined knowledge, attitudes, and behavior could be conveyed.

In the second half of the 20th century, the formation of the core qualities of the current educational system was more or less completed (Zymek, 2011). Since then, classrooms consist of 25 to 30 students receiving a compulsory education based on a state-defined curriculum under the guidance of one teacher. Also, social research started exploring the structural properties of schools and classrooms. According to Talcott Parsons (Parsons, 1959), the reason for students' misbehavior cannot be reduced to their lack of character and discipline or a failed upbringing. Instead, the structural nature of the social system of classrooms causes mental strain on students which will cause some students to disturb lessons. According to Parsons, schools have two

functions. For children, schools have the function to prepare them for their future roles in adult society. In schools, students internalize the "commitments and capacities for successful performance of their future adult roles" (Parsons, 1959, p. 297). However, for society, schools have the function to be the selective institution that "allocate[s] these human resources within the role-structure of society" (Parsons, 1959, p. 297). It is this often-overlooked selective function that sets schools apart from other agencies of socialization and causes student strain and as a result lesson disturbances.

To explain the selective function of schools, Parsons shifts his focus from the institutional level of the school to the teaching setting of the classroom. Within classrooms, students are selected by placing their achievement on a standardized continuum that divides them into two groups: Those students that will receive an academic education in college and those students for whom an academic education will be out of reach. This selective function and the way it is implemented in classroom settings can lead to strain for those students whose educational achievement leads to a selection in the non-college contingent. These children are regularly confronted with the fact that they are being selected to become part of the lower valued – in social status and wage – strata of society. For students, the teacher is the representative of school and adult society and therefore identified as the source of the strain. As a result, students dealing with strain do not only develop an indifference to classroom activities, but they can also show overt revolt against their teacher.

2.1.2 Classroom interaction.

Where social science from a structural-functional perspective focuses on the coercive effects of school structures, science with a focus on interaction uncovers some of the interaction mechanisms that make classroom management such a challenging task (Herzog, 2009). These mechanisms include firstly role-taking of both teachers and students and secondly the elements of classroom interaction.

Role-taking

In his seminal work *Life in Classrooms*, Philip W. Jackson (Jackson, 1990) describes some of the

main properties of classroom interaction and the role that students and teachers play in the social classroom setting. Teachers are expected to act as gatekeepers to classroom interaction. It is expected that he or she decides the topic, form, and participants of every classroom communication. A teacher is also a time-keeper that despite the "elaborate system of bells and buzzers" (Jackson, 1990, p. 12), still has the responsibility to meticulously keep track of time for each of the classroom activities and interactions. As a result, the average teacher performs over 1.000 student-teacher interactions per day, ranging from managing turn-taking, explaining a difficult topic, giving feedback, praising, and reprimanding. Additionally, a teacher constantly needs to make quick decisions with incomplete information (Borich, 2011).

Also for Jackson, the fact that students do not participate in schools of their own free will is of critical importance. In classrooms, students must learn to deal with three important factors that differ from other interactional settings, like for example the family: *crowds*, *praise,* and *power*. Firstly, the students' behavior is always in the presence of others, both peers and the teacher. Secondly, this behavior is subject to constant evaluation. Thirdly, students need to cope with the teacher as an authority figure that has much more power than the students. For Jackson, these factors lead to students developing adaptive strategies; the most important one is waiting. Teachers tend to put strong sanctions on student interactions that interfere with classroom order. Therefore, students tend to disengage temporarily from classroom interaction. They need to wait for their turn in a classroom discussion or recitation, when they want to ask a question, or when they have finished their assignment and need to wait for the others to complete. This waiting makes it a challenging task for children to stay engaged in learning and not turn to behavior that disturbs teaching. Moreover, to Jackson, the absence of disturbing behavior is a bad indicator of the effectiveness of a teaching measure. Although student engagement is a condition for learning, the amount of engagement in learning is often overestimated. "The signs of overt attention are not always trustworthy indicators of the pupil's actual state of mind" (Jackson, 1990, p. 102).

Elements of classroom interaction.

Where Jackson points to the interaction from a student's perspective, Walter Doyle points to the

difficulty of the classroom setting for the teacher. He defines six core elements of the classroom environment: *Multidimensionality*, *Simultaneity*, *Immediacy*, Unpredictability, *Publicness,* and *History* (Doyle, 1986). Firstly, *multidimensionality* refers to the fact that the people involved in classroom interactions, typically around 25 students and one teacher, have different and sometimes contrary needs and preferences, not necessarily all of which are related to teaching and learning. Moreover, these actors have a limited amount of recourses – time, attention – at their disposal to reach their goals. For example, teachers need to meet curriculum schedules, evaluate student-achievement, distribute working materials, set up technical equipment and communicate school-notifications. A student, on the other hand, might need teacher feedback on his or her work, further explanation on a difficult topic, or hand in an assignment. Moreover, a student might also want to discuss events from the weekend, previous lessons, or recess with a peer. This *multidimensionality* makes the classroom a place where many interests vie for the available time and attention.

Secondly, *simultaneity* refers to the fact that many of the classroom events described above occur at the same time. When a teacher turns to a specific student to answer her question during seatwork, he or she must coincidently scan the classroom for the possible questions of other students, keep track of time and monitor potential or actual lesson interruptions. This simultaneity increases when exercise forms and content are more heterogeneous.

Thirdly, *immediacy* refers to the fact that teachers do not have much time to reflect before acting. In most classroom situations, teachers are required to (re-)act immediately. To illustrate this immediacy, Doyle states that teachers "publicly evaluate pupil conduct with either praise or reprimands on the average of 15.59 times per hour" (Doyle, 1986, p. 394).

Fourthly, *unpredictability* refers to the classroom as a place where the chain of events cannot be anticipated. These events are produced through constant interaction between all participants. Therefore it is often difficult to anticipate how a planned activity will proceed on a particular day with a particular group of students.

Fifthly, *publicness* refers to the fact that all members of the group witness teacher-student interactions. A teacher's reprimanding of a student, for example, is observed and evaluated by

all other students. When a teacher fails to notice certain student behavior or unjustly reprimands a student, this is noted by the other students and might be a reason to join in on an interaction that initially only existed between a teacher and another student.

Sixthly, *history* refers to the fact that classes meet regularly over a long period. In the course of this period, the participants in the classroom setting develop a set of routines through this shared experience. The planning of a single event, therefore, must always take the history of the class involved into account.

In sum, the interaction perspective shows the social complexity of classroom settings and the respective differences in roles between students and teachers. Moreover, it shows that classroom order is not a result of the combined students' misbehavior. Rather, the heterogeneous and simultaneous interactions present in classrooms make it a big challenge for a teacher to manage classroom interaction in such a way that learning can take place.

2.1.3 Students' and teachers' perspectives on classroom order.

In their review of empirical studies regarding the perspectives on the topic of classroom management, Anita Woolfolk-Hoy & Carol S. Weinstein found that the beliefs and perceptions of students and teachers reflect their respective roles, objectives and relative power (Woolfolk Hoy & Weinstein, 2006). Firstly, student beliefs related to classroom management mainly focus on the subordinate role of being the object of socialization. Woolfolk-Hoy and Weinstein (ibid.) state that several empirical studies have shown that students consider a teacher to be good when he or she can establish a caring relationship with them. However, what 'care' means seems to differ between different groups of students. For high-achieving students ‘care’ is defined as assistance in academic matters. Low-achieving students tend to associate care with certain personality traits like patience, humor and the ability to listen to students.

Furthermore, Woolfolk-Hoy and Weinstein (ibid.) point out that a positive correlation exists between teachers who care about their students and the degree to which students are willing to be attentive and conscientious. This willingness to reciprocate is most evident for students with stigmatized or underprivileged backgrounds. Apart from establishing a caring relationship with

students, a teacher is also expected to manage classroom interaction. From the perspective of students, teachers are considered 'good' when they provide behavioral limits for students, keep classroom order and create a classroom environment where students feel safe. Accordingly, students are willing to engage more in lessons and have more respect for teachers who can exercise authority without being rigid, threatening or punitive.

Woolfolk-Hoy and Weinstein also found that regarding classroom management, teachers are mainly concerned with order and academic achievement (Woolfolk Hoy & Weinstein, 2006). Additionally, order and achievement are considered to be strongly connected. Teachers want their students to be cooperative in class and at the same time to perform well on assessments. Behavioral interventions are only considered effective when they increase cooperative behavior and academic achievement. Nonetheless, the degree to which achievement and management are considered related strongly depends on the competence of the teacher. "The most successful teachers view class management as the creation of effective, engaging, supportive learning environments and the socialization of students, whereas less successful teachers see classroom management as discipline and the maintenance of authority" (Woolfolk Hoy & Weinstein, 2006, p. 193).

These perspectives on classroom management confirm the classroom as a social setting where socialization and selection take place. Both students' and teachers' adapt their perspectives to their relative position in this socializing system.

2.1.4 The difficulties of managing classroom order.

The research discussed in this paragraph shows why classroom management is such a challenging task for teachers. Firstly, classrooms are places where students do not participate out of their own free will. Student disturbances cannot be reduced to their lack of character or failed upbringing. Rather, the socialization and selection function of a classroom might make a student feel coerced to disturb lessons. Secondly, classrooms are places in which the interaction is highly complex. This complexity makes it hard for a teacher to understand what is happening and also leaves very little time and space for teachers to find an appropriate strategy. Thirdly, students tend to be aware of their subordinate position in classes and tend to prefer teachers that show

care and attentiveness to student needs. At the same time, however, students expect teachers to set limits on certain student behavior.

2.2 Academic Perspectives on Classroom Management

This paragraph aims at providing an understanding of how teachers can effectively influence the challenging qualities of classroom interaction. For this, we take a closer look at the empirical findings of different schools of scientific research on the topic. Textbooks on the subject of classroom management categorize the distinct research perspectives differently (Haag & Streber, 2012; Mayr, 2006; Ophardt & Thiel 2008; Seidel, 2009). In this text, the discussion of these perspectives is placed according to three functions of classroom management: classroom management provides a method for (a) dealing with student disruptions, (b) managing classroom interaction and (c) increasing of available learning time. For this, we use the three academic approaches listed in the standard work by C.M. Evertson & C.S. Weinstein in their Handbook of Classroom Management (Evertson & Weinstein, 2006b): the *Behavioristic* approach, the *Ecological* approach, and the *Process-Outcome* approach.

2.2.1 Behavioristic approach.

The aim of the behavioristic approach lies in assessing and comparing the effectiveness of individual teacher interventions on student lesson disturbances. The main assumption is that student behavior can be influenced by applying stimulus-structure of punishment and praise (Ophardt & Thiel, 2013). Desired student behavior is encouraged by the application of stimuli that imply a personal reward or benefit for the student. Undesired student behavior is ended by stimuli that imply a form of punishment or the retraction of previously earned rewards. At the base of the behavioristic perspective lies the assumption that when correct stimuli are applied consequently, students will tend to show the desired behavior permanently. The role of research from a behavioristic perspective then lies in identifying those most effective stimuli.

Empirical research evaluates the effects of different behavioral stimuli by simulating specific classroom situations in laboratory settings (Landrum & Kauffman, 2006). The applied method is oriented on other research within the behaviorist tradition and typically consists of four phases.

In the first phase, the nature and cause of problematic student behavior are analyzed. The three following phases follow the steps of experimental research. In the implementation phase, a specific stimulus is started, and its effects are measured by the degree to which the problematic student behavior changes. In the retraction phase, the stimulus is halted and its effect on problematic student behavior assessed. In the last reimplementation phase, the stimulus is restarted, and it is assessed if the effects on problematic student behavior can be observed again. It is assumed that the changes in student behavior can be ascribed to a specific stimulus when the problematic student behavior is reduced upon application, reemerges upon retraction and is again reduced upon the renewed application of the stimulus.

Empirical research shows the effectiveness of several stimuli (Ophardt & Thiel, 2008: 265-266). Giving praise and paying attention had a strong effect within the experimental settings. Ignoring the undesired student behavior, a technique called *extinction*, was especially effective when it was combined with stimuli that encouraged desired student behavior. Also, the implementation of a so-called *response-cost-management* proved to be effective. The technique involves the retraction of rewards earned in the past.

This research also showed that many of the effects that could be observed in the experimental settings were much weaker when they were extrapolated to the field of real teaching or when they were generalized to other teaching subjects and students' age groups. This indicates that research following a pure behaviorist approach has difficulties taking in the complexity of real classroom interaction. These difficulties strongly limit the validity of the findings of research within the traditional behaviorist perspective. Especially under the influence of research within the ecological tradition, the understanding of classroom management as representing a set of interventions to deal with students' misbehavior is criticized.

Despite these criticisms, the main contribution of the behavioristic perspective lies in the formulation of techniques that help to deal with lesson disruptions that are assumed to be more or less generic. Indeed, textbooks concerning teachers' possibilities of dealing with the challenge of student discipline are based (at least partly) on behavioristic assumptions (see for example Canter, 2010). Current research within the behavioristic perspective typically divides these

techniques into two categories. The first category consists of those techniques that reinforce desired student behavior like giving praise and approval, introducing a system of earned points (for example for advantages in grading) and rewards or privileges, the establishment of clear rules and directions. The second category consists of those techniques that decrease undesired behavior like the extinction of a particular behavior, self-evaluation and self-reprimanding by students, time-out-measures, techniques for relaxation and the reduction of anxiety and the retraction of earned points and privileges (Brophy, 2006, p. 27).

2.2.2 Ecological approach.

The ecological approach states that the social processes of the classroom setting are more complex than assumed by the behavioristic approach. The six elements of the classroom environment (Doyle, 1986) discussed in chapter 2.1.2, explain why the effects of stimuli found in experimental settings cannot be replicated under the conditions of real teaching practice. This complexity can only be accounted for when teachers move their focus from the level of individual students to the level of the whole class (Doyle, 2006). In the ecological approach, classroom management focuses on managing classroom interaction in such a way that disturbances are prevented, and learning can take place.

The first main theoretical component is the analytical difference between students and the classroom. To ecological theorists, classroom teaching consists of the challenges of the enabling of learning and the keeping of order (Doyle, 1986, p. 395). Both challenges are closely related: a minimal degree of order is required for learning to take place, and learning activities must be structured in such a way that they capture student attention. However, the ecological approach implies an analytical separation between learning and order. "Because individuals rather than groups learn, an analysis of learning directs attention to individual processes. But order is a property of a social system and thus needs to be framed in a language of group processes" (Doyle, 1986, p. 395).

A *program of action* refers to the dominant activity structure (teacher's and students' interaction) that applies to a certain lesson segment. Ideally, the program of action encompasses a specific learning activity, like for example, seatwork or recitation. The program of action defines the

appropriate student behavior and the instructional and managing behavior of a teacher. (Hastie, 2006). Merritt (Merritt, 1982, p. 224) suggests that programs of action consist of *vectors of activity* that "once entered into, pull events and participants along their course" (Doyle, 2006, p. 102). These vectors of activity can be divided into primary and alternative vectors. Primary vectors refer to the objective that the teacher has for a specific lesson part. During the lesson, students develop secondary vectors compete with the primary vector for domination within a program of action.

Instead of viewing classroom order as the product of the sum of all individual deviations of classroom rules, order can only be analyzed on the level of the group. Therefore, order is seen as a construct that is negotiated socially between all participants of a certain group, between teacher and students and also between individual students. Following this argument, order is not the result of the degree to which student conform to behavioral rules. More so, it is dependent on students' willingness to cooperate with a teacher. (Doyle, 2006, p. 100). Secondly, classroom order must be seen within its specific social context. For example, the degree to which talking to a classmate is seen as an alternative vector of action depends on that specific context: the lesson phase, the learning activity and the rule-system of that specific teacher. Absolute silence can be considered indispensable, for example, while taking a test. However, classroom order does not imply absolute silence per se. A teacher's classroom management should be focused on strengthening students' preparedness to cooperate under the social and instructional context within a specific lesson situation. Order "means that within acceptable limits the students are following the program of action necessary for a particular classroom event to be realized in the situation" (Doyle, 2006, p. 99). Order, therefore, depends on the degree to which a teacher can define and consolidate the primary vector and shield it against secondary vectors.

For researchers working within the ecological approach, it is crucial to assess the context of the classroom setting in their full complexity. Therefore, typical research analyzes visual recordings of classroom interaction. In recent decades, classroom interaction was mostly recorded on video, but early studies were also done analyzing time-lapse photography (Gump, 1967). To identify the programs of action within classroom interaction research within the ecological perspective consists of two steps: dividing the chronicle of each recording into *segments* and identifying

activity structures for each of those segments. Segments "represent natural units of organized action" (Doyle, 2006, p. 101) and define the timely boundaries within a lesson chronicle. These timely boundaries are identified by looking for changes within (a) the patterns for the physical arranging of participants, (b) the use of props and learning materials, (c) the activity format, and (d) their corresponding rules of appropriate behavior (Dinkelaker & Herrle, 2009). As a next step, for each of the segments, the corresponding *activity structure* is defined. This means that firstly, for each segment, its *purpose* (i.e., testing, practice), *activity format* (i.e., lecture, recitation, discussion), and *topic* or *assignment* (language, math) must be defined. Secondly, the respective *roles* of the teacher (instructing, monitoring) and students (cooperating in activity format), as well as their interaction, are defined (Burns & Anderson, 1987, p. 36). The degree to which teacher's classroom management has been effective is expressed in quantifying student engagement. By using quantitative data-analysis, characteristics of teacher behavior are identified that have a positive influence on student engagement and therefore, classroom order.

The most influential study from the ecological approach was done by Jacob S. Kounin. Kounins (Kounin, 1970) seminal work consists of two parts. In the first part, Kounin assessed the so-called *ripple effect*: the effect that a teacher's disciplinary intervention aimed at a single student 'ripples out' towards other students within the group that are not targeted by the reprimand. Because the findings from these first studies were inconsistent, a second study was performed following the logic of the ecological approach. Kounin formulated six *dimensions* of classroom management. These dimensions refer to a quality of teacher behavior that shields the primary vector of action and therefore, positively influences classroom order. *Withitness* refers to the degree to which a teacher can communicate to students that he or she knows what is going on. To show that he or she has 'eyes in the back of his or her head.' Indications for a high degree of withitness are when the reprimands are timed well (not too late in the chain of interaction) and are focuses the right students. *Overlapping* refers to the degree to which a teacher can attend to several classroom interactions simultaneously. A teacher shows a high degree of overlapping when he or she does not need to interrupt the flow of the program of action in an attempt to end secondary vectors of action. *Smoothness* refers to a teacher making sure that the main program of action is not interrupted or changed and follows a clear objective. *Momentum* refers

to the teacher's ability to apply pacing of learning activities that keep the majority of students engaged and avoids slowdowns. *Group focus* refers to the ability of a teacher to focus on all students without losing sight of individual students. Group focus is achieved, for example by choosing formats of interaction and participation that keep all students engaged. Lastly, *satiation* refers to the ability to choose a broad variety of learning activities that do not last too long and are made interesting to students.

The main contribution of the ecological approach lies in broadening the scope of classroom management. Classroom management does not consist of a 'repertoire' of universally applicable techniques to deal with student interruptions. Instead, classroom management refers to behavioral qualities that create or sustain classroom order. These qualities cover a broad range of overt and more subtle teacher behavior that depends on the specific context of a specific lesson segment and can be categorized along the six dimensions formulated by Kounin.

2.2.3 Process-outcome approach.

Research within the process-outcome tradition focuses on identifying teaching factors that have a positive effect on student learning. Therefore, the main quality of the process-outcome approach is to shift the focus from student engagement to on the dependent variable of student achievement. Through the research within the process-outcome perspective, findings from research within other perspectives are integrated, specified, replicated, or put into perspective (Ophardt & Thiel, 2008, p. 271). For example, several studies could confirm the findings of Kounin regarding *withitness*, *overlapping,* and *smoothness*, but additionally show their positive effect on student achievement (Brophy, 2006, p. 29). As such, the process-outcome approach does not offer an alternative theoretical approach. Instead, it focuses mainly on integrating several findings from different perspectives and relates them to student learning. However, there are some important conceptual differences between the process-outcome perspective and the previously reviewed perspectives.

In their review article, Gettinger & Kohler (Gettinger & Kohler, 2006) introduce the process-outcome approach to classroom management by referring to the four main groups of variables relevant to research on effective teaching. The first group of variables is referred to as *presage*

variables. These variables refer to individual factors that each of the participants involved in teaching and learning brings into teaching setting like age, sex, social class, cognitive abilities, experience, attitudes, beliefs, and expectations. *Context* variables refer to the structural, social, and institutional influences on teaching like grade level, subject matter, instructional objectives, and class size. *Process* variables refer to the interaction happening within the classroom setting like teacher-student interaction, group management strategies, and learning activities. *Product* or *outcome* variables then refer to measurable indicators of student learning progress like grades, learning progress, or changes in student attitudes. The process-outcome approach of classroom management only focuses on the last two groups of variables, in which the first group constitutes independent variables, and the last group constitutes dependent variables.

Although these groups of variables might not be as comprehensive as more sophisticated models (Helmke, 2012; Slavin, 1994), they help to recognize that both classroom management strategies, as well as instructional variables, can fall under the scope of the research approach. Although research within the process-outcome approach separates both groups of process variables conceptually and empirically (Gettinger & Kohler, 2006), research has shown that both kinds of process variables are related. Firstly, because both variables affect student learning, but also because both groups of process variables influence each other. As stated in the paragraph discussing the ecological approach, classroom order can be increased by implying managing strategies and improvements in instruction. (Munk & Repp, 1994). In this section, however, we will only focus on the influence of classroom management variables.

Typical process-outcome research uses two different methodologies (Gettinger & Kohler, 2006, pp. 75–76). Firstly, low-inferent observational instruments are developed that assess a specific teacher behavior or classroom interaction that is hypothesized to be effective. Secondly, these instruments are then used to monitor and rate teacher behavior during a certain observation period. Additionally, student achievement is assessed, for example, through grades or knowledge tests, before and after this observation period. If the level of student achievement changes from pre-test to post-test, the correlation between observed classroom interaction as assessed by the instrument and changes in outcome are analyzed. Although process-outcome research also includes additional research methods, like for example, high-inference ratings or teacher

interviews, low-inference ratings are used more prevalently because they typically show the strongest correlation with student achievement.

An alternative methodology consists of the evaluation of teacher training programs in an experimental setup. In the first step, teacher-student interaction is observed to identify teaching behaviors that vary in frequency across classrooms and are expected to relate to student achievement. Then a sample of teachers is assigned to two groups: an experimental group of teachers that receives training in that effective teaching behavior and a control group does not receive training. The effectiveness of that training is evaluated by comparing teacher behavior and student achievement between the experimental and the control group.

Research in the process-outcome approach has strongly contributed to the establishment of an extensive knowledge base of the effectiveness of strategies for classroom management (Doyle, 1985; Duffy, 1990). Below, characteristics of effective classroom management strategies are summarized that have most often been replicated in process-outcome research (Gettinger & Kohler, 2006). Some of these strategies have already been discussed in the previous section on the ecological approach.

Firstly, a system of *classroom rules* makes sure that students know what is considered to be appropriate behavior for a certain lesson situation. The use of classroom rules is most effective when (a) rules are implemented and practiced systematically, (b) are implemented at the beginning of the year (Emmer, Evertson, & Anderson, 1980) and (c) their compliance is monitored and asserted consequently. Secondly, *learning time* should be used efficiently. This can be done by (a) reducing non-instructional activities and therefore maximizing the available learning time (Helmke, 2007) and by (b) minimizing off-task student behavior through the use of proactive classroom management strategies. Thirdly, teachers should make sure *transitions* between learning activities are smooth and quick, (Gump, 1982) because disruptive student behavior tends to increase significantly during unstructured transitions (Doyle, 1984). The effective management of transitions consists of (a) signaling the onset of transitions in advance, for example by using a certain sound or (b) actively planning and structuring transitions and (c) minimizing the loss of momentum. Fourthly, fixed *routines* or *procedures* allow a teacher to

initiate the instances of a lesson that tend to occur repeatedly without losing much time for instruction (Anderson, Evertson, & Emmer, 1980). Routines can be used both for the quickening of lesson transitions as well as for instruction or certain learning activities (Borich, 2011). When implemented effectively, routines make sure that no extensive instruction is needed each time a new lesson phase starts. Like a system of rules, routines should be systematically practiced, preferably at the beginning of the school year (Gettinger & Kohler, 2006).

There are two main kinds of criticisms of the findings from process-outcome perspective. Firstly, results should not be mistaken for a recipe book for all teacher behavior in all teaching situations. The effect on process variables represents general effects that are valid at an aggregated level. Secondly, the use of student achievement as a dependent variable does not take other important outcomes of teaching into account, like, for example, the creation of student learning attitudes and personal development. Despite these criticisms, findings from process-outcome research have made several important contributions to the field of research on classroom management. Firstly, and most notably, the research has shown the importance of classroom management not only for the sake of having orderly lessons but also for the degree to which students learn. Secondly, research within the process-outcome perspective has contributed strongly to the development of many standardized instruments that can be used to observe and assess teaching quality. Some instruments were developed for the use within the field of educational science, (Helmke & Renkl, 1992), others focus on the training and self-reflection of in-service teachers (Borich, 2011; Good & Brophy, 2008). Thirdly, the quantitative nature of the process-product research on classroom management allows it to be included into meta-studies. These meta-studies (Seidel & Shavelson, 2007; Wang et al., 1993) have shown the importance of effective classroom management, also in comparison to other aspects of teaching. This has led to the inclusion of classroom management competency into influential studies representing models of teacher competence (Blömeke, Kaiser, & Lehmann, 2010; Helmke & Jäger, 2002; Kunter et al., 2011). The main merit of research within the process-outcome perspective does not lie in the shift of the conceptual understanding of classroom management. Its main strength is to show the relevance of different classroom management strategies for student learning.

2.3 Conclusion

In this chapter, it was shown that students' disrupting behavior on classroom order is largely a result of the structural properties and functions of schools, the roles that teachers and students take in classroom interaction and the complexity of interaction within classroom settings. This helps to understand why situations, where teachers will have to deal with classroom disruptions, emerge in the first place and why finding effective classroom management strategies for teachers can be so challenging.

This chapter has also shown that research on classroom management can be grouped into three approaches: the behavioristic approach, the ecological approach, and the process-outcome approach. These approaches take on different perspectives on the definition of classroom management and which strategies are considered effective. The discussion of research on classroom management along the lines of these three approaches provides two advantages. Firstly, it allows for the historical, methodological, and theoretical categorization of the findings of the discussed studies within the broader research paradigms of educational a social science. Secondly, it helps to identify the academic roots of the broad range of different suggested classroom management strategies as they are proposed in textbooks (see the introduction to this chapter).

Regarding the theoretical assumptions, the three approaches differ strongly and, on some points, even take on contrary positions. However, looking at the results - what teachers can do so that lessons can proceed in an orderly manner and lesson disturbances are minimized, the approaches complement each other. This explains the heterogeneity of the different strategies. In their description of what classroom management is and what teachers can do to improve their management skills, these textbooks may refer to findings from all three approaches.

For the understanding and of classroom management, the current study does not focus on one of the three approaches. It is concluded that in each of the approaches, relevant empirical research was conducted. However, the current study does not place equal values on all three approaches. Especially the limitations of the behavioristic approach in its ability to generalize the effectiveness of intervention strategies should be regarded as a serious flaw. Despite this flaw,

the behavioristic approach helps to focus on the concrete teacher behavior that influences classroom order. Moreover, it reminds us, that classroom interruptions might occur, even when strategies related to the other approaches have been effectively applied. Teachers, therefore, should be equipped with interventional strategies to deal with these emerging interruptions. The ecological approach shows that classroom management strategies cannot be discussed without considering the complexity of social interaction in the classroom setting. The effectiveness of different strategies has been researched most systematically within the process-outcome approach. The most important contribution of research of the process-outcome approach is to understand why classroom order is favored in the first place. Effective classroom management leads the maximizing of the available learning time, increasing the probability that students learn something from the lesson. This discussion is summarized, using the the following foci on classroom management.

Focus on managing group interaction:

The ecological approach shows that a teacher's classroom management should focus on preventing disturbances of classroom order. Effective ways to preventing disturbances to develop are the management of group interaction, for example by increasing the degree of withitness and overlapping or by increasing monitoring possibilities for example through the effective use of the physical arrangement of the classroom.

Focus on increasing the effective use of available learning time:

Research from the process-outcome approach has shown that increasing the use of available learning time is one of the most effective ways to increase student achievement. Therefore, classroom management strategies should focus on increasing student engagement. The process-outcome approach also shows that the implementation of routines for teaching instances that occur regularly increases the use of available learning time. Moreover, the effective implementation of rules ensures that students know what behavior is expected.

Focus on managing lesson disturbances:

Even when group interaction is effectively managed and rules and routines have been effectively

put into place, there will be instances where a teacher will need to deal with students that disturb the lesson. Therefore, teachers should be able to deal with these kinds of situations. The behavioristic approach shows that several effective techniques can effectively end student disturbances. The ecological approach shows that these techniques should be focused on a minimal interruption of the primary vector of action. Therefore, teachers' reactions to student disturbances should be executed on the level with the lowest impact on lesson flow.

3. Classroom Management Competence

A professional's ability to skillfully solve specific professional challenges is typically expressed through the concept of competence (Klieme et al., 2008). However, many different understandings of competence exist, resulting in sometimes unclear discussions about the topic (Weinert, 2001). This chapter aims to answer the question, how prospective teachers' classroom management competence should be modeled so that, firstly, it can be validly measured and secondly, effective treatments to strengthen that specific competence can be developed. This chapter discusses three different models of teacher competence and their relevance for the teaching aspect of classroom management: *professional competence*, *professional vision,* and the *PID-Model*.

The chapter starts by providing a general overview of the concept of competence (3.1). Subsequently, an influential model of teacher competence, *professional competence*, its main components and conceptual structure are discussed (3.2). These components are applied to classroom management competence through the discussion of several empirical studies regarding knowledge (3.2.1) and beliefs on classroom management (3.2.2). Finally, several studies are discussed that call for a more situated approach to teacher competence (3.2.3). In the subsequent section, one of these situated models labeled *professional vision* is discussed (3.3). The section starts by describing the ethnographic origins of the concept (3.3.1). Then, the concepts application to research on teacher competence (3.3.2) and it use for standardized measurement of teacher competence (3.3.3) are discussed. An alternative situated approach to competence views competence as a continuum between the poles of a dispositional and performance-approach and was further worked out as the so-called *PID-Model* (3.4). The idea of viewing competence as a continuum is explained (3.4.1), and the core concepts of perception, interpretation, and decision-making are discussed (3.4.2). Expert-novice research regarding situation-specific skills on classroom management are presented (3.4.3), after which a section is dedicated to the concept of attentiveness that is expected to influence situation-specific skills (3.4.4). For the purpose of illustration, two research projects that have operationalized a situated approach to the modeling and measuring of classroom management competency are discussed

(3.5). The main theoretical assumptions and its operationalization within a video test are discussed regarding the projects VIU (3.5.1) and CME (3.5.2). The chapter is concluded with a summary of its main findings (3.6).

3.1 The Concept of Competence

The concept of competence refers to a relatively broad collection of different definitions and theoretical constructs (Klieme et al., 2008; Weinert, 2001). Despite these heterogeneous definitions, there are several assumptions that all concepts of competence share. Firstly, *competence* refers to a set of necessary mental prerequisites to successfully meet complex professional demands. Secondly, these prerequisites have cognitive as well as motivational, volitional, and social components. Thirdly, these components can principally be acquired through training or schooling (Weinert, 2001). In educational research, the concept of competence was initially used primarily for assessing and comparing the output of different educational systems on students' competence (Klieme et al., 2008; Shavelson, 2010) in large-scale studies like TALIS (Klieme & Vieluf, 2009) PIRLS (Mullis, Martin, Gonzalez, & Kennedy, 2003) and their 'mother studies' (Tillmann, 2016) TIMSS (Mullis, Martin, Foy, & Arora, 2012) and PISA. In these studies, a more narrow definition of competence is used that focuses on the cognitive elements of competence: competence is defined as "context-specific cognitive dispositions that are acquired by learning and needed to successfully cope with certain situations or tasks in specific domains" (Klieme et al., 2008, p. 9).

3.2 Professional Competence

Several large-scale assessments like MT21 (Blömeke, 2008), COACTIV (Kunter et al., 2011), TEDS-M (Blömeke et al., 2010) and LEK (König & Seifert, 2012) have applied the concept of competence to model and measure teachers' competence. The studies mentioned above developed similar models of teachers' competence that focus on the different mental components or dispositions of teachers' competence. One of the most influential models of teacher competence is labeled *professional competence* and was initiated through a text by Jürgen Baumert and Mareike Kunter (Baumert & Kunter, 2006). In this text, *professional competence* is positioned against the so-

called *structural approach*[1] that has mainly been influential in the German language area. According to proponents of the structural approach, professionals working within the pedagogical field are confronted with the apparent impossibility of combining the professional task of educating, socializing and schooling of children and the pedagogical task to meet an individual students' needs of care, security, and freedom (Helsper, 2004). A teacher is challenged with the "requirement of doing the one thing, without refraining from doing the other thing"[2] (Combe & Kolbe, 2008). These antinomies can then only be overcome by increasing the amount of case-knowledge, through hermeneutical analysis (Combe & Kolbe, 2008). The structural approach sees the teaching profession as characterized by deficits, lacking clear technologies, and impossible to professionalize (Tenorth, 2006). Baumert and Kunter criticize this "quasi-therapeutic" (Baumert & Kunter, 2006, p. 469) perspective on the relationship between teacher and student. Instead, they call for a shift away from this deficit-oriented approach and suggest a pragmatic competence-based approach that they label *professional competence*.

The model of professional competence (see Figure 1.) was developed as part of the COACTIV-study (Baumert & Kunter, 2011, 2013) and aims at identifying the competence aspects that teachers need to manage their professional challenges. COACTIV distinguishes between a generic structural model and a 'core' model (Baumert & Kunter, 2013). The generic model adopts a broad approach to competence. It consists of four aspects that influence teaching practice: values and beliefs, motivational orientations, self-regulation skills, and professional knowledge. The core model adopts a narrower cognitive approach to competence and focuses on this last aspect of professional knowledge.

[1] *Strukturtheoretischer* Ansatz, Translation in English by the author.

[2] Translation by the author.

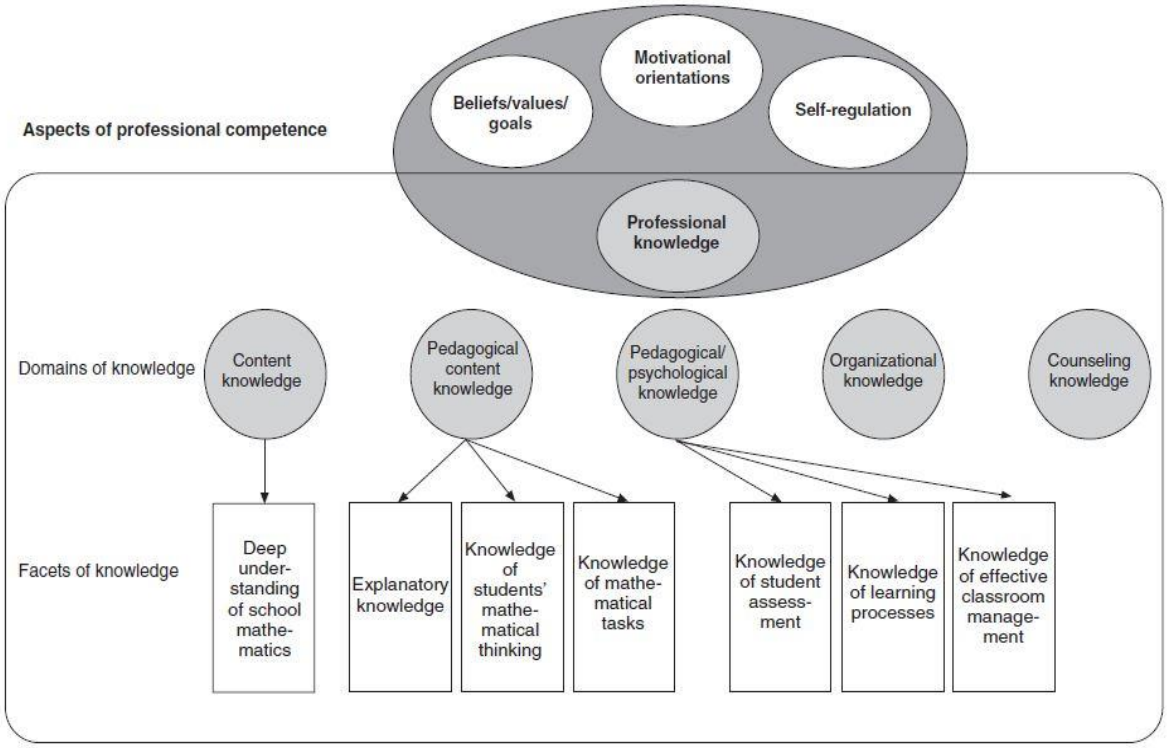


Figure 1. **Model of Professional Competence (Baumert & Kunter, 2013, p. 29)**

3.2.1 Knowledge of classroom management.

The focus of the model of COACTIV lies on the knowledge-aspect of competence, which is referred to as the "core of professionalism" (Baumert & Kunter, 2013, p. 28) and is structured as follows. Professional knowledge is assumed to consist of three *types* of knowledge. Conceptual or *declarative* knowledge consists of the collection of static facts, concepts, and principles that can be accessed by memory. *Procedural* knowledge refers to the knowledge on behavior of how to apply declarative knowledge in teaching practice. *Strategic* knowledge that refers to meta-cognitive knowledge-structures relateds to the knowledge regarding problem-solving strategies (Gruber & Stamouli, 2015).

Building on the seminal work by Lee S. Shulman (Shulman, 1986, 1987), each of these three knowledge-types falls in one of three knowledge-*domains*: (a) c*ontent* knowledge refers to knowledge on the teaching subject, (b) p*edagogical content* knowledge refers to knowledge on subject-specific teaching methodologies that are aimed to convey content knowledge in teaching

situations, (c) pedagogical-psychological or *general pedagogical* knowledge, refers to the generic subject-independent knowledge on learning and teaching (Voss, Kunina-Habenicht, Hoehne, & Kunter, 2015, p. 202). The separation in three knowledge types and three knowledge domains is also used within similar models of teacher competence, like for example the model used within the TEDS-M project (Blömeke, 2008). Baumert and Kunter (Baumert & Kunter, 2013) have added a fourth an a fifth knowledge-domain: (d) *organizational* knowledge refers to knowledge of the educational system and (e) *counseling* knowledge refers to the ability to communicate effectively with laypersons.

Within the model of professional competence, strategies for effective classroom management are located within the domain of *general pedagogical* knowledge (Blömeke & König, 2010; Voss, Kunter, & Baumert, 2011). The model of professional knowledge uses a broad understanding of classroom management, based on the process-outcome approach (see chapter 2.2.3). It assumes classroom management to consist of the following four facets: (a) knowledge on patterns of instructional practice, (b) knowledge on variation of social forms and methods of learning, (c) knowledge on rules and routines of effective classroom management, and (d) knowledge on creating a constructive and supportive learning environment (Baumert & Kunter, 2013, p. 35).

It could be shown that knowledge on classroom management is not of conceptual, but mainly of procedural nature (Cocard & Krähenbühl, 2013). In other words, teachers' relevant knowledge on classroom management concentrates on how to act in specific situations rather than focusing on abstract factual knowledge categorized along conceptual dimensions. Consequently, tests for the assessment of classroom management competence from the perspective of professional knowledge tend to assess procedural aspects of knowledge, for example asking test-participants to mention different effective classroom management strategies (König & Blömeke, 2009) or to assess their self-reported knowledge on classroom management strategies (Thiel, Ophardt, & Piwowar, 2013). As part of the COACTIV project, a video test was developed to measure procedural knowledge on classroom management (Voss, Kunter et al., 2011). Instead of asking test-participants to state their knowledge, this test measures the degree to which test-participants can describe and interpret a classroom management situation depicted in a short video. From the above mention studies, we can conclude that knowledge of classroom

management should be considered an aspect of competence consisting mainly of procedural knowledge and strongly dependent on the teaching context.

3.2.2 Beliefs and attitudes on classroom management.

In the models of COACTIV and similar research projects like TEDS-M (Blömeke et al., 2010), teacher beliefs are considered an important aspect of a teacher's professional competence. The concept of beliefs is sometimes referred to as a "messy construct" with multiple, partly contradictory definitions that are hard to "clean up" (Pajares, 1992, p. 329). Although beliefs and knowledge are assumed to be separate aspects of teacher competence, also within the model of professional competence, the separation between knowledge and beliefs is assumed to be blurry (Baumert & Kunter, 2013). One generally accepted definition understands teacher beliefs as the subjective assumptions that teachers make regarding the nature of a specific teaching aspect that influences a teacher's behavior (Blömeke, Suhl, & Döhrmann, 2012; Voss, Kleickmann, Kunter, & Hachfeld, 2011). Within the model of professional competence, beliefs are placed primarily in relation to the teaching subject (Blömeke et al., 2012; Dunekacke, Jenßen, Eilerts, & Blömeke, 2016), like for example beliefs on the nature of student learning, beliefs on the teaching subject, and beliefs on the effectiveness of teaching methods (Voss, Kleickmann et al., 2011).

The model of professional competence assumes that the distinction in three knowledge-domains, *content* knowledge, *pedagogical* content knowledge, and *general pedagogical* knowledge, not only applies to cognitive knowledge-types, but also to non-cognitive knowledge-aspects, like beliefs. As such, it is assumed that also beliefs relate to the subject, its teaching methods, but also to subject-independent teaching aspects like classroom management.

Several studies have shown the relevance of teacher beliefs for classroom management competence. According to Nancy Martin et al. (Martin & Yin, 2006; Martin, Yin, & Baldwin, 1998; Martin, Yin, & Mayall, 2007), teachers tend to have different classroom management styles that depends on that teacher's belief on the control of student behavior. This style of management is dependent on two dimensions of beliefs regarding classroom control: *instructional management* and *people management*. Beliefs related to instructional management refer to the maintenance of behavioral rules, the prevention of lesson disturbances, and the monitoring of student work.

Beliefs related to people management refer to the nature of students' individuality and autonomy, the degree of students' freedom during work, and how teachers should develop a fitting relationship with students. A teacher's management style is dependent on the degree of control on each of the dimensions that a teacher believes to be appropriate.

These dimensions are operationalized in a test called the *Attitudes and Beliefs on Classroom Control* (ABCC) inventory. This test creates a score for each of the dimensions of beliefs. An example item of the instructional management dimension is "I assign students to specific seats in the classroom." An example of the people management dimension is "I believe that students need direction in how to work together" (Martin et al., 1998, p. 8). Experienced teachers tended to show significantly higher degrees related to the control of instructional management but significantly lower degrees related to the control of people management (Martin & Yin, 2006). The authors conclude that this difference is the result of novice teachers having naive assumptions concerning 'automatic' cooperation between teachers and students. Expert teachers, on the other hand, despite having relatively low scores on controlling related to people management, still hold the belief that behavioral rules should be actively implemented and student work strictly monitored. This research shows that beliefs play an important role within the classroom management competence as beliefs differ between novice and expert teachers.

3.2.3 Situated models of teacher competence.

In recent years, several studies have criticized some of the assumptions of models of teacher competence like the model of professional competence. The criticism centers around the notion that models of teacher competence do not sufficiently take the context of the teaching-setting into account (Borko 2004; Cochran-Smith, Feiman-Nemser, & McIntyre, 2008; Oser, 2013; Oser, Heinzer, & Salzmann, 2010). These studies argue that models of teacher competence that only focus on representing the mental dispositions (knowledge, beliefs, and self-regulation), do not sufficiently predict performance in practice (Blömeke, Busse, Kaiser, König, & Suhl, 2016; Shavelson, 2010). Reliable measurement of complex constructs like competence implies the necessity to break up the construct into its constituent dispositions, each of which can then be reliably measured. However, this dispositional approach bears the risk that the sum of those

reliably measured individual dispositions cannot validly depict the complexity of the construct of in situ teaching performance. In other words, measuring competence leads to a reliability-validity dilemma (König, 2015a).

Another criticism focuses on the risk of conceptual ambiguity, as formulated by Hans Neuweg (Neuweg, 2011). According to Neuweg, there are three types of meaning of the concept of teacher knowledge that are often used indiscriminately. Firstly, knowledge can refer to the collection of systematic declarative knowledge that is codified within the curriculum of programs of teacher training and acquired at universities. Secondly, knowledge can refer to a psychological construct, consisting of the cognitive structures of teachers that can be assessed by psychometric research instruments. Thirdly, knowledge can refer to the dispositions that are used to describe or explain teacher behavior in practice. This last type of knowledge implies that when a teacher 'knows' something, for example, how to manage classrooms, this means that he or she can manage classrooms in the daily teaching settings. Failing to differentiate between these three kinds of knowledge is problematic in two ways. Firstly, declarative knowledge acquired in teacher training (type 1) is not automatically transferred into practice (type 3) (Mägdefrau & Schumacher, 2001) and therefore runs the risk of remaining tacit (Neuweg, 2004). Secondly, it is criticized that the design of paper-pencil knowledge-tests (type 2 knowledge) is based too much on the curricula of programs for teacher training (type 1 knowledge), while falsely assuming to be automatically valid for teaching practice (type 3 knowledge) (Lüders, 2012).

Following these criticisms, several researchers have called for competence models that focus less on the assessment of knowledge. Instead, it is called for the development and operationalization of models of teacher competence that are better suited to take the complex reality and context of teaching into account (Oser, 2013; Cochran-Smith et al., 2008). A promising method for this situated measurement of competence is seen in the use of video-based standardized tests (Seidel & Thiel, 2017). Two influential models are discussed in the following sections: *professional vision* and the *PID-Model*.

3.3 Professional Vision

One of the initiatives to find more situated approaches to the modeling and assessment of

teaching competence is labeled *professional vision*.

3.3.1 Ethnographic origins.

The concept of professional vision was first introduced by the anthropologist Charles Goodwin (Goodwin, 1994) in an article in which he shows how groups of professionals reproduce (archaeologists) and content (police officers and lawyers) their respective profession through the use and analysis of visual materials. In his words, professional vision "consists of socially organized ways of seeing and understanding events that are answerable to the distinctive interests of a particular social group." (Goodwin, 1994, p. 606). Goodwin makes clear that groups of professionals define, demarcate, reproduce, and contest their respective professional field through the use of socially constructed mechanisms for working with visual materials. For the profession of archeology, for example, Goodwin mentions that "[t]he ability to see in the very complex perceptual field [is] central to what it means to see the world as an archaeologist. Such seeing would be expected of any competent archaeologist." (Goodwin, 2001, p. 174).

3.3.2 Application on teacher competence.

The concept of professional vision was introduced to the field of educational research by Miriam Gamoran Sherin (Sherin, 2001). Her work is mainly based on observations of mathematics teachers' oral reflections within so-called 'video-clubs' in which groups of teachers discuss video clips of their own teaching. Sherin and colleagues have developed an understanding of professional vision that describes and classifies teachers' interpretation of visual materials (Sherin, 2007a; Sherin & Han, 2004; Sherin & Van Es, 2009; Van Es & Sherin, 2008). Professional vision consists of two separate but connected and simultaneous mechanisms: selective attention or *noticing*, that refers to the interactions that teachers in video clubs focus on, and *knowledge-based reasoning* that refers to the cognitive level of teachers' reasoning about what was noticed in the video. It was found that during a video club, teachers firstly tend to focus more on pedagogical issues (what are the alternatives for pedagogical strategies could the teacher have implemented), and secondly more on student thinking (why does a student have difficulty understanding an explanation on a certain mathematical problem). Regarding knowledge-based reasoning, during a video-club, teachers tend to move from mainly *describing* student behavior

to the higher cognitive levels of *evaluating* the importance of the student's statements and *interpreting* the meanings of student ideas. This understanding of professional vision moves away from the ethnographic description of qualitative differences between different professional groups (Lefstein & Snell, 2011). Instead, it presents professional vision as a way that a certain level of teachers' competence manifests itself. In other words, a teacher's level of competence can be determined by looking at what a teacher *notices* and by looking at what level of *knowledge-based learning* he or she uses.

3.3.3 Standardized measurement.

In recent years, several research projects have taken up this understanding of the concept of professional vision and developed standardized instruments. One of the leading research projects is called *Observe* (Jahn, Prenzel, Stürmer, & Seidel, 2011; Seidel et al., 2010; Seidel & Stürmer, 2014; Stürmer & Seidel, 2017). In this project, a video test called *Observer* was developed that is capable of reliably and validly measuring prospective and in-service teachers' noticing and knowledge-based reasoning on topics within the domain of pedagogical-psychological knowledge. *Noticing* is considered to describe "whether teachers pay attention to events that are of importance for teaching and learning in classrooms." (Seidel & Stürmer, 2014, p. 4). Based on meta-analysis (Seidel & Stürmer, 2014), the following three components for noticing are defined: (a) *goal clarity and orientation*, (b) *teacher support and guidance*, and (c) *learning climate*. Knowledge-based reasoning is considered to exist on the levels of (a) *description*, (b) *explanation,* and (c) *prediction*. The video test uses short video clips as item-prompts (Kersting, 2008) to answer rating-items. A participant's competence is expressed by the percentage of agreement with a master-rating. Through several validation studies, the factorial structure of Observer regarding the three levels of knowledge-based reasoning could be confirmed with a sample of prospective teachers (Jahn, Stürmer, Seidel, & Prenzel, 2014). Also, it could be shown that Observer is sensitive to the measurement of changes in professional vision resulting from video-based courses in university (Stürmer, Seidel, & Schäfer, 2013). Observer has been used in several studies aiming to use a standardized situated measure for the assessment of competence within the domain of general pedagogical knowledge (Mertens & Gräsel, 2017;

Seidel & Stürmer, 2014; Stürmer & Seidel, 2015).

3.4 The PID-Model

An alternative model of situated competence was introduced by Sigrid Blömeke, Jan-Eric Gustafsson & Richard J. Shavelson (Blömeke et al., 2015b) (see Figure 2.).

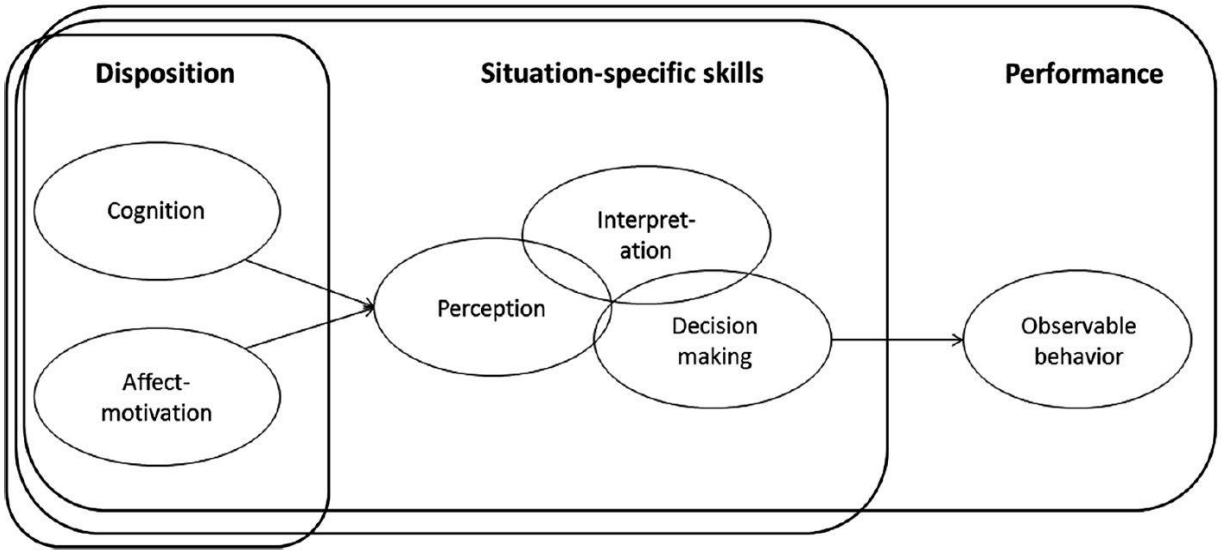


Figure 2. **PID-Model (Blömeke et al., 2015b, p. 5)**

These authors suggest taking an alternative perspective on the understanding of the concept of competence. They argue that the typical understanding of competence, defined as "complex ability constructs that are context-specific [...] and related to real-life" (Shavelson, 2010, p. 3), is too limited and has led the academic debate to be polarized between a dispositional and a behavioral approach to competence. Researchers favoring a *dispositional* approach focus on the investigation of the inter-relations between the different cognitive and affective-motivational competence aspects. Researchers favoring a *behavioral* approach focus on the assessment of observable professional behavior in real life situations. Because adopting either a dispositional approach or the behavioral approach leads to several conceptual and methodological problems, Blömeke et al. (Blömeke et al., 2015b) suggest that an alternative approach to understanding the concept of competence is needed. Competence should be seen as a continuum, in which mental dispositions and real-life behavior represent the two poles and so-called s*ituation-specific-skills*

form the middle ground. In the following sections, the PID-Model and its implications for the modeling of classroom management competence are further explained.

3.4.1 Competence as a continuum.

Within the *dispositional* approach, competence is viewed as a collection of specific, measurable latent traits, each of which should be measured reliably and validly. The validity of these measurements is then expressed by testing to which degree these measurements successfully predict performance in a real-life situation. In other words, in the dispositional perspective, practice is the criterion variable with which the measurement of the latent traits is validated. The origins of this approach stem from educational research and its objective to find effective ways to advance the development of a person's competence (Blömeke et al., 2015b). The main assumption is that to increase a person's competency, the most important latent traits underlying this competency should be identified so that they can be specifically trained in educational settings.

The *behavioral* approach, on the other hand, views competence as the observable behavior in real-life praxis. According to this approach, real-life practice is the constituent of competence itself. As a consequence, competence can only be measured firstly through the sampling of representative real-world tasks and secondly observing and evaluating performance within these tasks. In this holistic approach on competence, performance and its underlying mental dispositions are assumed to be indivisibly linked together. The relative weight of each disposition cannot be generalized, because it depends strongly on the context. Therefore, separating these dispositions through operationalization for the sake of reliable measurement leads to a serious reduction of validity. The origin of this approach lies in organizational psychology that aims at identifying suitable candidates within the recruitment process.

Blömeke et al. (Blömeke et al., 2015b) explicitly state that they polarize the characteristics of both positions. However, they describe the academic debate to be characterized by both positions "mutually criticiz[ing] each other fiercely for misconceiving the construct [of competence], reducing its complexity, [and] ignoring important aspects" and that "the value added by each of the perspectives is rarely acknowledged" (Blömeke et al., 2015b, p. 5). From a

methodological point of view, the dispositional approach has several advantages. Most importantly, because many instruments rely on paper-pencil tests, measurements can be obtained with relatively low costs. Consequently, samples can be big, and instruments can be highly standardized and extensively piloted within large samples, leading to high degrees of measurement reliability. However, there are serious concerns regarding the validity of the representation of the target construct. The behavioral approach takes on a holistic view and assumes that the whole of real-life competence is greater than the sum of its dispositional constituent parts. From a methodological point of view, the behavioral approach has two disadvantages in comparison to the dispositional approach: reliability and representativeness. Observations need to be done in complex, messy, real-life situations. This means that measurements tend to be more time- and cost-intensive and a strong limitation of the number of observations exists, making it more challenging to reach the same values of measurement reliability. Additionally, because real-life behavior tends to be heavily situation-dependent, it becomes challenging to sample real-world situations so that they ensure representativeness for the whole range of professional challenges.

To solve these problems, Blömeke et al. (Blömeke et al., 2015b)) take a different conceptual understanding of competence. Instead of seeing the two perspectives as contrary and dichotomous, competence should be viewed as a continuum on which both perspectives are placed on either side. This is done, firstly, by identifying the theoretical assumptions that both approaches agree on: (a) competence is regarded as acquirable and can thus be improved by deliberate and specific practice or training and (b) real-world situations are considered important, either as a criterion variable or as the constituent of competence itself. Secondly, the central assumptions of both perspectives should be acknowledged: (a) mental dispositions can be individually identified and assessed and form the basis for performance within real-life situations; (b) these dispositions are strongly and systematically linked, are activated in response to in-situ professional demands, and their relative weight can change during the course of the performance.

The central question then becomes: how “persons who possess all of the resources belonging to a competence construct can integrate them, such that the underlying competence emerges in

performance?" (Blömeke et al., 2015b, p. 6). In other words: which competence aspects connect both ends of the continuum? To empirically answer this question would require the separate measurement of each possible individual disposition. Instead, the authors suggest a conceptual heuristic: so-called *situation-specific skills,* mental processes that are triggered through visual representations of real-life tasks. These situation-specific skills are *perception*, *interpretation,* and *decision-making*.

This new approach allows for the conceptual integration of both the dispositional and behavioral poles. This perspective on competence allows for the combination of each of the conceptual and methodological advantages while avoiding their respective disadvantages. Blömeke et al. (Blömeke et al., 2015b) expect these advantages to be especially valid for the professional fields in which a theory-practice gap in both training and assessment is traditionally heavily debated. One of these fields for which this is expected to be the case is teacher training. For the assessment of situation-specific skills, the authors see an important role for video-based assessments. Videos combine the advantage of both ends of the continuum. On the one hand, they can portray representative professional situations in their full complexity. On the other hand, they allow for standardization, which keeps costs for assessments limited.

3.4.2 Situation-specific skills.

Several academic initiatives have further developed this alternative approach to competence for the professional field of teaching (Mason, 2016). The initial heuristic was labeled as the *PID-Model, r*eferring to the first letters of the three situation-specific skills, perception, interpretation, and decision-making, (Kaiser, Busse, Hoth, König, & Blömeke, 2015). Before the initiation of this PID-model, a considerable number of studies had already investigated the perceptive, analytical and sense-making strategies of teachers when watching video clips (Stahnke et al., 2016). Therefore, one of the main tasks lies in the allocation of these studies into the framework of the PID-model.

However, this posterior step makes it challenging to define a precise understanding of the individual situation-specific skills. Indeed, to the knowledge of the author, none of the studies referring to the PID-model have yet provided any precise definitions. Although the different

previous studies make use of concepts that all point in the similar theoretical direction, they do not necessarily mean the same thing. For example, during the course of the work of Elisabeth van Es and Miriam Sherin (Sherin, Jacobs, & Philipp, 2011a; Sherin & Van Es, 2005; Van Es & Sherin, 2002), the understanding of *noticing* seems to shift away from a purely perceptive concept towards a concept that includes both perceptive and reasoning elements. In one of the most recent texts on the topic, they state that "these two aspects of noticing are interrelated and cyclical. Teachers select and ignore based on their sense-making; the ways they respond shapes subsequent instructional events, resulting in a new and varied set of experiences from which teachers attend and make sense" (Sherin, Jacobs, & Philipp, 2011b, p. 5).

In the following sections, the situation-specific skills *perception* and *interpretation* are discussed based on findings from these exploratory expert-novice studies. Additionally, theoretical and empirical considerations regarding *decision-making* are presented.

Perception

Based on extensive research on teachers watching videos of mathematics lessons in primary schools, Frederick Erickson (Erickson et al., 1986; Erickson, 2011a) summarizes the main features of teacher *perception* or what he calls *noticing* of video recordings. Firstly, *noticing* refers to the filter through which teachers select relevant situations within the complex whole of the teaching practice: teachers attend to some phenomena and ignore others. Secondly, noticing is instrumentally opportunistic: teachers notice things of which they think teacher action or interventions are required. Thirdly, noticing is influenced strongly by teacher's attitudes on pedagogy and teaching. Fourthly, noticing is strongly influenced by the amount of a teacher's experience: prospective teachers can notice accurately, albeit in fragmentary ways. Novices have difficulties linking those events in their causal relationship and tend to discern only isolated situations. In contrast, expert teachers can meaningfully connect these separate occurrences. For example, they express accurate assumptions regarding the period of the year and the institutional context. Also, expert teachers can integrate different interactions, behavioral details, and classroom characteristics into one 'story frame', linking the connections between individual events in a cause-effect relationship.

These findings are confirmed and complemented by other studies. A comparison of expert and novice teachers' behavior watching series of photographic slides showed that novices made shorter and more generic descriptions about static features of the classroom, indicating that they can remember fewer details and are not able to infer relationships between several events (Carter et al., 1988). Moreover, Copeland et al. (Copeland, Birmingham, DeMeulle, D'Emidio-Caston, & Natal, 1994) found that experts were better able to identify causal relationships between teacher and student actions than novices. A study by Star & Strickland (Star & Strickland, 2008) and its replication study (Star, Lynch, & Perova, 2011) explored what prospective teachers focused on when watching the full-length video in an eighth-grade mathematics classroom before and after training on teaching methodology. Both studies found that before participating in the video analysis course, prospective teachers possess weak skills in observing classrooms. The strongest effect of the intervention firstly lies in noticing surface properties like the physical *classroom environment:* estimates of the number of students or the presence of an overhead projector, and secondly with the dimension of *classroom communication:* how teachers gave directions and responded to student comments.

Interpretation

Willis Copeland (Copeland et al., 1994) investigated the differences between expert and novice teachers' interpretation of videos. Firstly, experts were better able to relate events seen in the video clips to theoretical knowledge of teaching and learning. Secondly, experts were able to infer the goals related to student learning from the observed teaching seen in the video clips. Thirdly, experts related the events seen in the video clips more to events and students that were not directly portrayed in the video clip. They also made suggestions about what should be added or changed in that lesson. Fourthly, experts tended to relate the perceived situations to issues of student learning, whereas novices seemed to be more interested in issues related to behavioral control and classroom management.

These findings were confirmed by several other exploratory studies. The investigations of Sherin et al. (Sherin & Van Es, 2009), for example, confirmed that with the increase of expertise, teachers relate their observations more to inferences regarding student thinking. This last point

was also found in the study of Carter et al. (Carter et al., 1988). It could be shown that the level of interpretation or *knowledge-based-reasoning* could move from mere *description* to the higher levels of *evaluation* and *interpretation* through participation in video clubs for in-service teachers (Sherin & Van Es, 2009) and video-based courses for prospective teachers (Van Es & Sherin, 2002).

Decision-Making

Of all three situation-specific skills, decision-making is the least included in studies using video tests. Also, it is the skill that is least developed in prospective teachers (Stahnke et al., 2016). Decision-making is also considered to be the situation-specific skill that is located closest to practical performance on the continuum of competence (see Figure 2.). Therefore, it is assumed that decision-making is the situation-specific skill that novice prospective teachers possess the least.

Some writers use a broad definition of decision-making, which, for example, also includes the *proposing of improvements* (Bischoff, Brühwiler, & Baer, 2005; Santagata & Yeh, 2016). Blömeke et al. (Blömeke et al., 2015b), however, seem to propose a more narrow understanding based on the work of Alan Schoenfeld (Schoenfeld, 2012, 2015). Schoenfeld uses several examples of teaching situations to point out how teachers make decisions. Thereby, he shows that teachers tend to make in-the-moment goal-oriented decisions based on knowledge, problem-solving strategies, metacognition, and beliefs. Schoenfeld's understanding of decision-making aims at modeling teachers' teaching practice, rather than the measurement of competence and the differentiation between novices and experts. However, by adapting Schoenfeld's understanding of decision-making, Blömeke et al. (Blömeke et al., 2015b) state that decision-making should be considered as a situated-specific skill that is based on *immediacy* (Doyle, 1986), time-pressure, and the need to decide without the possibility of extensive reflection. Also, it becomes clear that in-the-moment decision making is assumed to be strongly dependent on beliefs, for example on students' learning or teaching strategies.

No empirical studies are known that include this factor of time-pressure. However, several studies show that the simulation of the decision-making process is an indication of teaching

competence. Bischoff et al. (Bischoff et al., 2005) confirmed that experienced teachers were better able to formulate mathematical diagnostic interventions when prompted by a video clip. Bruckmaier et al. (Bruckmaier, Krauss, Blum, & Leiss, 2016) showed that the ability to formulate valid teaching decisions as prompted by a video clip positively correlates with pedagogical content knowledge on mathematics and constructivist beliefs on teaching. Moreover, test-scores positively correlated with students' ratings on teachers' quality of teaching, also indicating the relevance of the situation-specific skill of decision-making for teaching practice.

3.4.3 Situation-specific skills on classroom management.

Expert-novice comparisons that focus on classroom management competence confirm the above-described findings on general teaching for the specific field of classroom management. Regarding perception, the studies of Star et al. (Star et al., 2011; Star & Strickland, 2008) show that novice teachers have a strong tendency to focus on issues regarding classroom management. This was confirmed by a study of Sabers et al. (Sabers et al., 1991), in which it was shown that novices and advanced beginner focus much more on disturbing student behavior, even at the cost of events related to instruction and student learning. Moreover, novices "seemed to want to express their disapproval of students who wasted time, who paid little attention to what the teacher said or did, and who displayed a general disinterest in what was happening." (Sabers et al., 1991, p. 79). Experts, on the other hand, were not as harsh on student's disruptions as novices and advanced beginners. Also, experts tended to mention possible reasons for the student disturbing behavior. The study of Carter et al. (Carter et al., 1988) showed that when asked to stop the series of slides and focus on classroom management aspects, both experts and novices stopped at the same slides to make comments. However, there were differences in what was then discussed: experts consistently agreed between each other about which features of a teacher's behavior were salient about classroom management, whereas novice teachers contradicted each other or even themselves in their remarks.

Regarding interpretation, the studies of Charlotte Wolff et al. (Wolff, Jarodzka & Boshuizen, 2017; Wolff et al., 2016; Wolff, et al., 2015) provide a good insight into differences between expert and novice teachers. These studies compare novice and expert in-service teachers' verbal reflections

on aspects of classroom management within video-recordings of their own teaching (Wolff et al., 2015) and unknown teachers (Wolff, Jarodzka & Boshuizen, 2017). Experts firstly tend to provide significantly more predictions of intentions or feelings of students and expectations about student or teacher behavior. Expert teachers tend to relate specific events more to typical situations, showing their knowledge of similar cases. Also, expert teachers significantly placed a specific event into its temporal context, referring to other events in the video or the probability of events happening outside of the temporal boundary of the video clip. Finally, whereas novices related classroom management events to students being on- or off-task, experts related a classroom-management event to issues related to student learning. However, no studies were known that place the skill of decision-making within the context of classroom management competence.

3.5 Two Studies on the Situated Assessment of Classroom Management Competence

For the purpose of illustration and future reference, two models of situation-specific skills on classroom management competence and their respective methods operationalization and measurement are discussed: the project *VIU-Early Science* (Gold & Holodynski, 2017; Holodynski et al., 2017; Steffensky, Gold, Holdynski, & Möller, 2015) and a test named *Classroom Management Expertise (CME)* (König, 2015b; König & Lebens, 2012). These studies are described in detail as examples of the situated assessment of classroom management competence through the use of video tests.

3.5.1 VIU: Professional vision of classroom management.

In the VIU project, classroom management is understood to consist of those teaching strategies that lead to a positive learning climate, and that maintain a high rate of student engagement (Gold & Holodynski, 2017; Holodynski et al., 2017). Referring to research from the ecological (Doyle, 2006; Kounin, 1970) and process-product perspective (Evertson & Emmer, 2009), classroom management is assumed to focus on three aspects: *monitoring* of student behavior, *managing momentum,* and the establishment of *rules and routines*. *Monitoring* refers to the dimensions of withitness and overlapping, as formulated by Kounin (Kounin, 1970). This includes

a teacher being aware of all relevant interactions and a teacher's immediate and appropriate response to student disturbances. *Managing momentum* refers to the organizing of smooth transitions between classroom activities and clear instructions on desired learning behavior. *Rules and routines* refer to sets of behavioral expectations that are shared by students and teachers and therefore, are important for providing structure and orientation.

The video test is based on the concept of professional vision as developed by Tina Seidel and colleagues (Seidel et al., 2010; Seidel & Stürmer, 2014). The model discerns between *noticing* and *interpretation*. *Noticing* refers to the ability to filter relevant from irrelevant information. Through *interpretation*, teachers make sense of what is noticed. Interpretation consists of the abilities to explain and *review* the observed events and *predict* possible effects (Steffensky et al., 2015). Professional vision is assumed to be the result of knowledge, schemata, and scripts, but does not represent a simple categorization of knowledge (Seidel & Stürmer, 2014), because it also includes, beliefs, goals and cultural background (Steffensky et al., 2015). Professional vision of classroom management then is defined as "the ability to notice and to interpret those classroom events that require teachers to monitor, manage momentum and apply rules and routines" (Steffensky et al., 2015, p. 355).

The video test consists of four video clips, lasting between two and four minutes depicting different learning activities, such as teacher-class dialogue, seatwork, teacher-instruction, and lesson transitions. Twenty-nine experienced teachers rated these videos as authentic and fitting for the analysis of classroom management (Hellermann, Gold, & Holodynski, 2015). Forty-seven closed rating-items were developed: 18 items represented the facet of *monitoring*, 12 items represented the *managing of momentum* and 17 items represented *rules and routines*. Items referring to *noticing* aim at the recognizing of relevant aspects of classroom management. Items referring to *interpretation* aim at the correct inferring or conclusion of a noticed event. Test-participants rate their agreement to the items on a four-point scale. By way of example, Figure 3. shows some of the items that were used in the assessment of professional vision on classroom management (Steffensky et al., 2015, p. 360). Item-scores were created by comparing the ratings of the test-participants to a master rating created by a group of experts.

	Agree	Partly agree	Partly disagree	Disagree
Monitoring				
The Teacher reprimanded a student, although another one was, in fact, disturbing the class more *(noticing)*	□	□	□	□
Some students were distracted because the teacher was not able to prevent all interruptions quickly enough *(interpretation)*	□	□	□	□
Managing Momentum				
The transition from individual work to joint work was smooth *(noticing)*	□	□	□	□
Some students could not follow the lesson, because the teacher explained too quickly *(interpretation)*	□	□	□	□
Rules and Routines				
The teacher reminds students about the rules for good behavior in the lesson (Noticing)	□	□	□	□
The teacher has successfully established classroom rules (interpretation)	□	□	□	□

Figure 3. **Item-Examples of VIU Video Test (Steffensky et al., 2015, p. 360)**

Several studies validated the model of professional vision of classroom management as measured by the video test (Gold & Holodynski, 2017; Hellermann et al., 2015; Steffensky et al., 2015). Values of internal consistency were good for the test on the whole (Pretest α = .87, posttest α = .90), and the sub-scales at pretest (*monitoring* α = .71; *managing momentum* α = .61; *rules and routines* α = .71) and at posttest (*monitoring* α = .81; *managing momentum* α = .77; *rules and routines* α = .81) (Hellermann et al., 2015).

Confirmatory factor analysis in a validation study showed that all sub-scales of the test are highly correlated, but that a one-dimensional model, scaling the test by combining all items into one score fit the data best (Gold & Holodynski, 2017). Another study (Hellermann et al., 2015) confirmed that the video test is sensitive to differences in professional expertise. Evaluation of a university-based course on classroom management showed medium effect sizes for prospective teachers participating in their master's program (d = 0.34) and for prospective teachers participating in their masters and bachelors program (d = 0.54). A third study (Steffensky et al., 2015) confirmed that professional vision could be empirically separated into a subject independent construct (classroom management) and into a subject related (mathematical learning support) construct. A sample of 116 master-program prospective teachers and 125 in-

service primary school teachers participated in the above-discussed video test on *classroom management* and a similarly designed video test on *mathematical learning support.* Factor analysis showed that the constructs of professional vision of classroom management and professional vision of learning support are positively associated, but represent distinct constructs.

3.5.2 CME: Classroom Management Expertise.

CME assumes classroom management to be mostly focused on the managing of group processes that focus on the prevention of classroom disturbances like the practicing of classroom rules, setting consequences, monitoring, and the increasing instructional clarity (Casale, Strauß, Hennemann, & König, 2016). This understanding of classroom management is operationalized in CME through the situations depicted by four video clips. In each of these video clips, a teacher is challenged to manage "(1) transitions between lesson phases, (2) instructional time, (3) control student behavior, and (4) organize instructional feedback. (König, 2015b).

CME is designed around the three cognitive challenges *Accuracy of Perception* (AP), *Holistic Perception* (HP), and *Justification of Action* (JA). (König & Lebens, 2012). Firstly, expert-novice research shows that because of their categorical perception (Bromme, 1992), experts can meaningfully identify relevant instructional situations more precisely than novices (Carter et al., 1988) and can remember more relevant details and in greater detail. Novices, on the other hand, are not able to filter the complexity of the educational situation and therefore, are overwhelmed by a large amount of simultaneously occurring interactions in educational settings (Sabers et al., 1991). Consequently, novices have difficulties remembering relevant facts, because they lose themselves in the number of details, both relevant and irrelevant. This first cognitive challenge that is used as a measure of classroom management expertise is labeled *accuracy of perception*.

Secondly, novices tend to observe instructional situations as a collection of individual occurrences. Experts, however, can classify an observed situation as an example of typical teaching scenarios or curriculum script (Putnam, 1987), have an intuitive grasp of the situation and their knowledge is highly interlinked. Therefore, experts can observe classroom situations more holistically (Bromme, 2001) and infer the meaning of an instance within a larger chain of

interactions. As such, experts can reconstruct the instructional setting from individual situational and interactional occurrences (C. Wolff et al., 2015). This *holistic perception* (HP) allows experts to predict the future course of an observed classroom situation, deduce probable causes for the development of the observed classroom situation, and reconstruct alternative effective teacher actions.

Thirdly, experts can interpret the function of instructional events and sequences and can reason about the instructional intention and rationale of a specific action (Berliner, 1992). Novices, on the other hand, only seem to be able to describe the action itself without being able to reconstruct the *justification of action.* It is expected that this last cognitive demand is rarely needed in the day-to-day school life of teachers. However, it is assumed that this can be accessed and verbalized from long-term memory. Where holistic perception is conceived as a perceptive-representational process, justification of action is thought to be a process of interpretation.

CME consists of four video clips that last between 1:09 and 2:43 minutes that depict authentic teaching situations in which teachers manage challenging classroom management situations. CME consists of 24 items, of which five are multiple choice and 19 are open-response. Of these 24 items, *accuracy of perception* is assessed by 14 items, *holistic perception* is assessed by six items and *justification of action* by four items. Figure 4. shows four items and their respective cognitive demand. Test takers have seven minutes to work on all the questions related to each video clip. After a short introduction, test takers watch the first video clip on a centrally placed video screen without the possibility to pause or rewind. After the first video, test-takers are prompted to work on the seven questions of the first video. Items are presented to test-takers in a booklet. After the seven minutes run out, the next video clip starts immediately.

Please watch the video clip and answer the questions as best as you can.

1. Which function does the seating arrangement of students have for the lesson shown? (AP)
2. When does the situation displayed in the video take place? (HP)
   - □ A. At the beginning of the lesson (e.g., during the first 5 minutes)
   - □ B. During the first third of the lesson
   - □ C. During the last third of the lesson
   - □ D. At the end of the lesson (e.g., During the last 5 minutes)
3. Please name four different techniques employed by the teacher to gain her students' attention. (AP)
4. What happens immediately after the video sequence? (AP)

Figure 4. **Item-Examples of CME Video Test (König & Lebens, 2012, p. 14)[3]**

A coding rubric was designed to code all 19 open-response questions. Coding is done through several dichotomous criteria that are developed through "a complex and extensive interplay of deductive (from [the] theoretical framework) and inductive approaches (from empirical teacher responses)" (König, 2015b, p. 7). The coding rubric for the open-response questions consists of one or more low-inferent criteria per item. For the 19 open-response items, 48 criteria were developed. All test-entries are coded using these criteria, either coding with a '1' if the criterion is met or with a '0' if the criterion is not met by the statement of the test participant. These codes on the level of the criteria are recoded in a second round of coding on the level of the question. These codes then represent the item-scores.

CME was validated over the course of four studies. In a first study (König, 2015b), CME was administered to a sample of 119 in-service teachers with different levels of expertise. The current study shows internal consistency of α = .70 and an average item-total correlation of r_{it} = .36. In the same study, confirmatory factor analysis showed CME to be a one-dimensional construct (König, 2015b). A different study showed a significant medium-sized correlation between CME and a knowledge test for general pedagogical knowledge (König, Blömeke, Paine, Schmidt, & Hsieh, 2011; König & Kramer, 2015). Also, it was confirmed that in-service teachers outperform prospective teachers on CME and that prospective teacher participating in a trainee program

[3] English translations from German originals of questions 1 to 3 are taken from the supplementary Material of König and Kramer (2015). English translation from German original of question 4 was done by the author.

outperform prospective teachers from the first phase of teacher training. Moreover, it was shown that CME's test-score was a stronger predictor for instructional quality as rated by students than general pedagogical knowledge, confirming CMEs context-sensibility. Criterion validity was shown by a small negative significant correlation between CME and two burn-out scales. However, no significant correlation was found between CME and a scale for self-efficacy and one-third burn-out scale (König & Rothland, 2016).

3.6 Information-Processing and Attentiveness

As we have seen in the previous sections, situation-specific skills are typically measured through the use of video tests. As such, it can be expected that the capacity to process and structure visually complex information plays an important role. Indeed, many video tests implement a form of the ability to recall information from a video clip as an indication of *perception* or *noticing*.

Empirical studies have shown that the level of visual load plays an important role in the precision with which visual information can be successfully stored in short-term memory. Using three television monitors simultaneously, Sabers, Cushing and Berliner (Sabers et al., 1991) asked groups of experts, advanced beginners and novices to answer questions regarding issues of classroom management and regarding student instruction as depicted in the video clips. Results show that experts in comparison to novices and advanced beginners were less likely to suffer from visual and auditory overload. Experts were able to monitor and interpret parallel events in more detail, they could focus on the information presented on more than one monitor and they could simultaneously attend sound and vision related to the teacher's instructions. Willis Copeland (Copeland, 1987) argued that a teachers ability for *withitness* and *overlap* (Kounin, 1970) depends on a teacher' cognitive abilities of *vigilance* and *multiple attention.* His study showed that students of teachers who performed better in a video game that was based on the operationalizing these cognitive abilities were less likely to be off-task.

These studies indicate that the degree to which a teacher can deal with complex visual material influences his or her monitoring ability and therefore forms an indicator for his or her classroom management competence. Moreover, these studies indicate that a teacher's vigilance is not only an effect of experience or training. Also, relatively stable cognitive person traits, related to the

ability to process visual stimuli, influence the degree to which a teacher is able to monitor classroom management.

In psychology, the ability to successfully process stimuli from the sensory registers and store information in short-term memory is labeled *attentiveness* (Hasselhorn & Gold, 2013). Additionally, attentiveness is assumed to represent the cognitive ability to filter relevant from irrelevant visual stimuli (Schmidt-Atzert & Amelang, 2012). Empirical research shows that the filtering of relevant and irrelevant information within situations with a high load of visual information plays a role in successful classroom management (Sabers et al., 1991). Other studies indicate that this ability can be regarded as a relatively stable cognitive trait (Copeland, 1987). No studies were known that investigate the relationship between the performance on video tests and a teacher's degree of attentiveness. The above-discussed research, however, indicates that this could be the case.

3.7 Self-Efficacy in Classroom Management Competence

Perceived self-efficacy is refers to "the beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments" (Bandura, 1997, p. 3). Perceived self-efficacy is therefore thought to influence how people "feel, think, motivate themselves, and behave [and it includes] cognitive, motivational, affective and selection processes" (Bandura, 1994). Self-efficacy is assumed to be positively influenced by four factors: (a) a set of personal successful or *mastery* experiences, (b) *vicarious* experiences based on seeing other people successfully performing similar tasks, (c) social persuasion related to the positive feedback a person receives from others, and (d) the reducing of stress (Bandura, 1994; Schwarzer & Warner 2011). Although self-efficacy does not represent an aspect of teachers' competence per se, several recent empirical studies have shown its relevance for the understanding of teachers' classroom management and the role it plays in the transfer of knowledge between programmes of teacher education and the teaching practice.

Several empirical studies have explored the relation between teachers' perceived self-efficacy and their classroom management behavior (Dicke et al., 2014; Gold et al., 2017; Pfitzner-Eden, Thiel, & Horsley, 2014; Thiel et al., 2013). Consequently, teachers' self-efficacy is regarded

relevant for the understanding of classroom management competence. However, teachers' self-efficacy consists of self-evaluations rather than the objective measurement of teaching performance. As such it is more strongly based on confidence and resilience to setbacks, rather than on actual performance (Bandura, 1994). Consequently, the construct is generally regarded as relatively independent of competence (Schmidt-Atzert & Amelang, 2012) and is not included in the competence-models discussed in this chapter.

However, recent studies imply that self-efficacy plays an important role in the understanding of classroom management competence in another way. It was shown that high degrees of teachers' self-efficacy in classroom management are negatively correlated to teachers' rates of burn-out (Brouwers & Tomic 1999) and emotional exhaustion (Dicke et al., 2014). Additionally, and more importantly, recent research shows that only when levels of emotional exhaustion and burn-out are low, does starting teachers' level of pedagogical knowledge positively correlate to their classroom management performance (Seiz, Voss & Kunter 2015). This suggests that low levels of emotional exhaustion pose a condition for the transfer of pedagogical knowledge into teaching practice. Therefore, although not part of models of competence per se, it seems self-efficacy on classroom management could play a mitigation role, as it influences levels of emotional exhaustion and burn-out that can hinder knowledge transfer of starting teachers.

Empirical studies on the relation between situation-specific skills in classroom management and self-efficacy on classroom management show inconsistent results. As already reported in the previous section, no significant correlation could be found between CME and scales for self-efficacy (König, 2015b). However, a study using the video test developed within the VIU-Project reports a significant positive correlation with a small effect-size between self-efficacy and scores on the video test (Gold et al., 2017). Additionally, the study showed that scores on self-efficacy on classroom management improved under the influence of video-based courses for prospective teachers. Another study could confirm the positive effect of video-based courses on perceived self-efficacy in classroom management (Kumschick et al., 2017).

It is therefore concluded that prospective teachers' self-efficacy on classroom management is related to situation-specific skills on classroom management: whereas situation-specific skills on

classroom management are regarded as an objective measurement, self-efficacy on classroom management is regarded as being more closely related to self-confidence. More importantly however, self-efficacy in classroom management plays a mitigating role, as it is positively correlated to burn-out and emotional exhaustion which can hinder the transfer of conveyed knowledge into the teaching practice.

3.8 Conclusion

In the introduction to this chapter, the question was posed, how to model classroom management competence so that it can be validly measured and effectively strengthened. In this last section, this question is answered by summarizing and comparing both situated models that were discussed in this chapter: professional competence (3.8.1), professional vision (3.8.2), and the PID-model (3.8.3). The final section (3.8.4) summarizes the role of self-efficacy within this model.

3.8.1 Professional competence.

The model of professional competence as laid out by Baumert and Kunter (Baumert & Kunter, 2006) is considered to take a dispositional approach towards teacher competence. The model allows for the operationalization and reliable measurement of a broad variety of competence aspects relevant to teaching. The model has been widely applied for the measurement of a multitude of subject-related and subject-independent aspects of teacher competence (Baumert et al., 2011; Blömeke & König, 2010; König & Seifert, 2012). Although the model includes non-cognitive competence aspects, the model's primary focus lies on teachers' knowledge (Baumert & Kunter, 2013). This model makes clear that for classroom management competence declarative knowledge has only limited relevance (Cocard & Krähenbühl, 2013). Instead, classroom management competence is mainly based on procedural knowledge (Voss, Kunter et al., 2011) and beliefs (Martin et al., 2007).

The research discussed in chapter two indicates that effective classroom management is strongly dependent on the context of the classroom setting. This context is characterized by its multidimensionality, simultaneity, immediacy, unpredictability, publicness, and history (Doyle,

1986, see also chapter 2.1). These findings indicate that classroom management competence should be considered strongly dependent on the teaching context. Therefore, the criticisms that were formulated on the model of professional competence (see chapter 3.2.3) are especially applicable for classroom management competence. Consequently, for the modeling of classroom management competence, a more situated model than the dispositional approach to competence is preferred.

3.8.2 Professional vision.

Professional vision is one of the two models of teacher competence that takes on a more situated approach. One of its main strengths is that it is based on the results of ethnographic and qualitative research (Goodwin, 1994; Sherin, 2001). Indeed, the research from the project Observe related to general pedagogical knowledge (Seidel et al., 2010) and project VIU for the concept of classroom management (Steffensky et al., 2015) show the relevance of the concept of professional vision for the research on the ability to analyze visual representations of teaching. Moreover, results of project VIU (see chapter 3.5.1) have shown that the concept is applicable for the competence aspect of classroom management and is able to discern between novice and expert teachers.

However, in comparison to the other two models discussed in this chapter, the model of professional vision is relatively limited in its scope. Neither conceptually nor empirically, is the concept of professional vision systematically linked to other competence aspects like on the one hand knowledge, beliefs, motivation, and on the other hand in-situ performance. Therefore, it remains unclear, how this concept should be conceptually related to these other competence aspects. The concept of professional vision remains a relatively isolated aspect of competence that is equivalent to the ability to analyze video clips on teaching. For the modeling of classroom management of the current study, a different situated model of competence is used.

3.8.3 PID-model.

The PID-model views competence as a continuum and arranges different competence aspects on this continuum between mental dispositions and performance (Blömeke et al., 2015b). One of

the main advantages of the PID-model is its integrating character. Firstly, the PID-Model is assumed to be an overarching concept that integrates studies and concepts that were thus far theoretically unrelated, like among others studies on procedural knowledge (Voss, Kunter et al., 2011), professional vision (Seidel et al., 2010; Sherin, 2007a) and ability to analyze (Santagata, Zannoni, & Stigler, 2007). In their review study, Stahnke et al. conclude that a "considerable body of research contributions dealing with perception, interpretation, and decision-making (...) already exists" (Stahnke et al., 2016, p. 24). Although referring to partly different concepts, these studies typically refer to the same exploratory expert-novice research on the use of video clips, indicating that they operationalize similar concepts. Also, the PID-Model conceptually integrates the dispositional, situated, and performance aspects of competence. Although no large-scale studies or longitudinal research have tested these assumptions, individual smaller studies confirm the dependency of situation-specific skills on beliefs (Bruckmaier et al., 2016; Dunekacke et al., 2016) and declarative knowledge (Kaiser et al., 2015; König et al., 2014) on the one hand and teaching performance on the other hand (N. B. Kersting, Givvin, Thompson, Santagata, & Stigler, 2012).

The three levels of knowledge-based reasoning, description, explanation and prediction, seem to be largely equivalent to perception, interpretation and decision-making respectively. Therefore, the theoretical dimensions of the model of professional vision can be integrated within the PID-model. Since the PID model offers a theoretical framework that connects situation-specific skills to other competence aspects, the PID-Model seems to be the more encompassing of the two situated competence models. However, the assumptions that the situation-specific skills are equivalent to the three levels of knowledge-based reasoning have not yet been tested empirically. As the model of professional vision and its main instrument have been validated more extensively than the PID-model (Jahn, Prenzel, Stürmer, & Seidel, 2011; Seidel et al., 2010; Seidel & Stürmer, 2014; Stürmer & Seidel, 2017), professional vision is still a valid a model for the measurement of situated competence.

The model of professional competence is viewed as complementary to the PID model as the different dispositional competence aspects from professional competence can be placed in relation to teacher performance on the other end of the continuum and situation-specific skills

in the middle. Only the knowledge type of procedural knowledge is not easily assigned a place on the continuum. Some studies operationalize procedural knowledge on classroom management as the degree to which participants are able to formulate different classroom management strategies (König & Blömeke, 2009) others claiming to measure procedural knowledge make use of video tests (Voss, Kunter et al., 2011).

In the current study, it is argued that the degree to which the models of professional competence and the PID-model are complementary depends on what is understood as procedural knowledge. A crucial part of the PID-model is that situation-specific skills are strongly influenced both by knowledge and beliefs. In the model of professional competence, procedural knowledge is viewed as a cognitive disposition that is relatively independent of beliefs (see Figure 1.). This means that when procedural knowledge is operationalized as the ability to mention different classroom management strategies (König & Blömeke, 2009), it can be assumed that this competence aspect falls within the dispositional perspective of competence. If, however, procedural knowledge is alternatively operationalized as the ability to perceive and interpret a classroom setting in a video clip (Voss, Kunter et al., 2011), teacher's beliefs are assumed to play a role and it is argued that the test should be regarded as measuring situation-specific skills instead.

There are also several disadvantages to the PID model. On a theoretical level, posterior reconceptualization, for example by equating the concept of *noticing* from the model of professional vision to *perception* from the PID-model, can be theoretically misleading. Stahnke et al. (Stahnke et al., 2016) acknowledge this by stating that "different terms [are] used for the same aspect as well as the same terms [are] used for different aspects" (Stahnke et al., 2016, p. 24). Although much empirical research has shown that factors that can be labeled as situation-specific skills represent a valid way to measure teachers' competence, the exact boundaries between these situation-specific skills remain unclear.

Equating the different studies is also problematic on a methodological level. To exactly understand to what degree concepts like noticing, perception, knowledge-based reasoning, interpretation, and decision-making overlap, one must thoroughly compare the different ways in

which these concepts are operationalized (Neuweg, 2015). As many studies for pragmatic reasons report only superficially on the development of video tests, it becomes largely impossible to decide where the conceptual boundaries between the different concepts should be drawn. This point is confirmed by the two discussed studies related to situation-specific skills on classroom management. CME (König, 2015b) should be regarded as one-dimensional, indicating the three cognitive challenges used for operationalization, *accuracy of perception*, *holistic perception*, and *justification of action* represent largely the same construct. Validation of the video test from project VIU shows that it can also be considered one-dimensional (Gold & Holodynski, 2017).

Moreover, there are several studies that indicate that the performance on a video test depends on relatively stable personality traits related to the ability to process visual stimuli (Copeland, 1987). This would mean that the variance in scores on video tests are at least partly the result of personality traits rather than that they represent differences in teaching competence.

3.8.4 Self-efficacy in classroom management.

Discussion on the construct of self-efficacy in classroom management showed its relevance to the topic of classroom management competence. On the one hand, it should not be considered an aspect of teachers' competence as are, for example, declarative or procedural knowledge. Therefore, the construct is not included in the model of classroom management competence used in the current study. However, on the other hand, the construct has high relevance for the current study because of its mediation effect for the transfer of knowledge into practice for beginning teachers.

3.8.5 The model of classroom management competence.

It is concluded that the PID-model is the preferred model for the situational measurement of prospective teachers' classroom management competence in the current study. Firstly, the model allows taking in the complexity of the teaching situation. As shown in chapter 2, effective classroom management does not consist of a universally applicable recipe-like list of interventions and is therefore not reducible to a set of declarative knowledge. Rather, classroom

management strategies are strongly dependent on the complex classroom situation. Secondly, because of its integrating character, the PID-model allows for the conceptual integration of and building on empirical findings of the other models discussed in this chapter: On the one hand, dispositional perspectives on competence like the model of professional competence and on the other hand, to already existing concepts like the model of professional vision. However, the PID-model should still be considered a heuristic. Firstly, it remains unclear to what degree situation-specific skills represent competence aspects that can be empirically distinguished. Secondly, the predictive validity of situation-specific skills for in situ performance has not been sufficiently attested.

Summing up, in the current study, classroom management competence is considered to refer to the ability to notice and analyze video clips of classroom management situations. This ability is, firstly, assumed to be based on beliefs and mainly procedural knowledge on classroom management and is, secondly, predictive of classroom management in real-life teaching situations.

4. Acquiring Classroom Management Competence

This chapter aims to provide an overview of methods with which prospective teachers acquire classroom management competence. The chapter, firstly, discusses general research-based guidelines for the development of university-based courses on classroom management (4.1). The effects of participation in teaching internships are systematically discussed (4.2) by pointing to the specific effects on non-cognitive competence aspects, general pedagogical knowledge, and situation-specific skills. The specific mechanisms of and conditions for learning through the analysis of video recorded cases (4.3) are discussed by reviewing the method of video-based case-analysis and describing several specific guidelines for the development of video-based courses. In this section, a special focus lies on the choice for an instructional strategy. Firstly, the logic of the cognitive instructional strategy and its relation to cognitive load theory are discussed. Secondly, the alternative situated instructional strategy and its relation to situated learning theory are discussed. Then four empirical studies that have systematically compared both instructional strategies are reviewed. In the concluding section (4.4) the main arguments of the chapter are summarized and their relevance for the central theme of the current study is explained.

4.1 University Programs on Classroom Management

In her review article on courses on classroom management in US-based universities, Lara Stough (Stough, 2006) notes that only a minority of the US-based universities offer courses that are explicitly dedicated to that topic. This is mainly due to the limited role of the topic of classroom management within government-set standards for teacher education. A study evaluating the implementation of education standards (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, 2004) in the curricula of universities within the German state of North-Rhine Westphalia (Lohmann et al., 2011) suggests that, similarly, up to that point the topic of classroom management has not represented a priority within university-based curricula in North-Rhine Westphalia. However, a Delphi-study showed that different groups of German teacher trainers place classroom management as one of the ten most important topics of teacher training (Kunina-Habenicht et al., 2012).

Many studies have shown that university-based courses can positively influence classroom management competence of prospective teachers (Blüthmann, Ophardt, Thiel, & Felsberger, 2011; O'Neill & Stephenson, 2012). In a review article, Vern Jones (Jones, 2006) described the three most important evidence-based guidelines for the design of courses of prospective teachers. Presenting a strong diversity of interventions might overwhelm prospective teachers and give the impression that classroom management consists of a set of techniques. Therefore, firstly, regarding conveyed content, courses should focus on a limited set of general strategies. Secondly, programs should focus on offering many possibilities to practice, reflect, and implement the strategies, instead of focusing on conveying theoretical concepts. Thirdly, university-based courses on classroom management should focus on attitudes and beliefs. Educators should assist prospective teachers in examining and reflecting on their beliefs and assist them in clarifying misconceptions.

Moreover, Jones (Jones, 2006) underlines the importance of the integration of course-work and practical teaching experiences. This integration can be reached by providing prospective teachers with the opportunity to reflect on their own teaching experiences. Secondly, classroom management courses should provide prospective teachers with the possibility to reflect on cases taken from real teaching practice. This is mainly done through the analysis of video-recordings of real teaching situations. The impact of internships and other field experiences should be increased by regular visits of university-based supervisors and the implementation of structures of apprenticeship in schools. The following paragraphs take a closer look at both methodologies: video-based case analysis and internships.

4.2 Acquiring Teaching Competence through Internships

As mentioned in chapter 4.1, it is often assumed that classroom management competence can be acquired by participating in internships and other field experiences. Indeed, all the involved stakeholders, like prospective teachers, school-based mentors, policy-makers, and teacher educators, tend to have high expectations of the effects of participation in internships (Holtz, 2014). These expectations include an increase in reflection-competence, increase of motivation for the teaching profession and the bridging of the gap between theory and praxis.

However, in an article reviewing several decades of empirical research on the effects of internship-participation, Tina Hascher tempers some of these expectations (Hascher, 2012b). Firstly, most empirical studies are fragmented and often rely on subjective self-evaluations, rather than objective measurements. Secondly, internships can also have undesired effects, like the decrease of motivation and self-efficacy. Moreover, prospective teachers often do not critically reflect their aptitude for the teaching profession. Thirdly, inter-subjective differences tend to be big, because they are strongly dependent on the quality of the mentoring structures and institutional linkage between school and the university. One of the main reasons for this diversity of effects is that internships represent an unspecific and unsystematic intervention. Studies indicate major differences between internship programs regarding the number of lessons that internships teach autonomously (Gröschner et al., 2015) and big differences between the quality of mentoring (Hascher, 2006). Notwithstanding these big differences between individual internships, this section reviews some of the more recent research on the effects of participation in an internship.

In the following chapters, the word 'internship' refers to parts of the university-based curriculum in which prospective teachers acquire teaching experiences directly in a school. During internships, prospective teachers are typically encouraged to apply the knowledge acquired in university during autonomous teaching and, at the same time, try out what school life looks and feels like (Hascher 2012b). Students in these practical phases are typically guided by a university-based teacher, either through preparation courses or courses that are held in parallel. Internships, therefore, are a part of programs of formal education. Therefore, internships should firstly be discerned from teaching experiences that prospective teachers acquire external to educational structures, for example, as part of a side job as a substitute teacher. Secondly, internships should be discerned from the so-called 'Referendariat', the practical phase following the university-based education, and that is typical for German-speaking countries. The Referendariat can be understood as a state-led traineeship. Passing the tests of the Referendariat is a condition for application for a teaching job in German-speaking countries.

4.2.1 Effects of internships.

No empirical studies were found that specifically evaluate the effectiveness of internships on classroom management competence. Therefore, the studies discussed below relate to effects on a general pedagogical competence.

Effects on beliefs

During an internship, prospective teachers tend to be mainly concerned with getting appreciation from their school mentor (Hascher, 2012b, p. 113). This means that during their internships, prospective teachers tend to develop a strong focus on adapting the aims, beliefs, and teaching style of that specific mentor. This focus on the mentor goes at the expense of a focus on implementing learning objectives that are provided by universities, like for example the development of reflective attitudes on the learning process, the exploration of mechanisms of student learning, or the application of university-acquired knowledge (Hascher, 2012b). Moreover, one of the effects of the internships seems to lie in a reduced motivation for participating in courses of teacher training programs (Holtz, 2014). However, Tina Hascher (Hascher, 2012b) concludes that one of the main effects of participation in internships is that prospective teachers have a more realistic notion of what it means to teach, what school-life looks like, and what the role of the teacher within classrooms is.

Effects on general pedagogical knowledge

A longitudinal study showed an increase of general pedagogical knowledge during the course of programs for teacher training (König, 2013). Moreover, it was shown that participation in an internship leads to a small significant increase in general pedagogical knowledge (König & Seifert, 2012). More specifically, Johannes König and Sigrid Blömeke (König & Blömeke, 2012) showed that the general pedagogical knowledge of 1,501 American and 1,032 German prospective teachers significantly correlated to the amount of independent teaching and the extent of mentor support received during their internships.

Effects on self-efficacy

Scales of self-efficacy are often used to measure the effects of internships (Hascher 2012b, Hascher 2006). Several studies confirmed the effectiveness of the participation of internships on self-evaluations of competence. A study questioning 169 prospective teachers before and after a five-week internship showed significant increases in all four sub-scales of self-evaluations of situated competence ranging from a middle effect-size (d = .41) to a strong effect-size (d = .90) (Gröschner & Schmidt, 2012). A second study evaluated the effects of the longer so-called practical semester (Gröschner, Schmitt, & Seidel, 2013). The practical semester is a relatively new internship-format for German programs of teacher training. Prospective teachers are immersed in schools for a semester, during which they regularly - usually once every one or two weeks - attend university-based seminars in which general pedagogy, teaching methodology, and prospective teachers' practical experiences are reflected. Using the same instrument as the one in the first study, 236 prospective teachers reported their subjective competence before and after participation in the practical semester. The study confirmed that prospective teachers rated their situated competence significantly higher on all four sub-scales after participation in the practical semester.

Another study could confirm these findings for the effects of the practical semester (Mertens & Gräsel, 2017). The scales developed by Gröschner (Gröschner & Schmidt, 2012) and a test on declarative general pedagogical knowledge were used to compare an experimental group of 142 prospective teachers participating in the practical semester to a control group of 35 prospective teachers not participating in an internship. The analysis showed that in the experimental group, increased their self-efficacy on pedagogical knowledge significantly higher than in the control group. However, the effects of internships on general pedagogical knowledge, as reported by an earlier study (König & Blömeke, 2012), could not be confirmed: no significant differences between the experimental and control group were found.

Effects on situation-specific skills

To the knowledge of the author, only the above-discussed study reports on the effects of internships on situation-specific skills (Mertens & Gräsel, 2017). Using the instrument Observer

(Seidel et al. 2010) this study found a small significant increase of participation of the practical semester on professional vision (η^2 = .06).

4.2.2 Conclusion: acquiring competence through Internships.

The empirical studies discussed in the previous sections show the differential effects of internships on prospective teachers' competence: situation-specific skills, self-efficacy (Gröschner & Schmidt, 2012) and an objective video test (Mertens & Gräsel, 2017) show significant increases as a result of participation in the practical semester. Regarding declarative knowledge, the evidence is less definite: although several cross-sectional and longitudinal studies showed positive effects regarding general pedagogical knowledge (König, 2013; König & Blömeke, 2012; König & Seifert, 2012), these results could not be reproduced in a quasi-experimental pretest-posttest study (Mertens & Gräsel, 2017).

Regarding the non-cognitive aspects of competence, research summarized by Tina Hascher (Hascher, 2012b), shows that participation in an internship leads to a more realistic notion of school life and teaching with prospective teachers (see chapter 4.2). It can be concluded that the effects of participation in an internship firstly lie in the changing of attitudes and subjective theories about teaching and secondly on the field of the situation-specific skills regarding general pedagogical knowledge.

4.3 Acquiring Competence through Video-based Case-Analysis

Around the turn of the millennium, educational researchers started to discover the use of video-based case analysis within programs for teacher training (Krammer & Reusser, 2004). Up to that point, the advantages of the medium of video had already been known for several decades (Erickson, 2011b), but had mainly been used for the scientific analysis of teaching itself. For many teacher trainers, these advantages also apply to teacher-training courses (Krammer & Reusser, 2004). Firstly, videos provide a 'window' into authentic teaching. Secondly, video recordings allow learners to explicate, reflect upon, and discuss their subjective theories. Thirdly, video recordings facilitate the translation and application of theories into teaching practice. Fourthly, videos allow a change of perspectives, for example, between a subject-oriented and a

pedagogical-oriented perspective (Helmke & Helmke, 2004).

4.3.1 Case-analysis.

In recent years, video recordings are increasingly used in programs for teacher training (Petko, Haab & Reusser, 2003), in many cases, through online platforms (Petko, Prasse & Reusser, 2014). The aim of the analysis of video-based cases in programs for teacher training is to foster some form of situation-specific skills (Steffensky & Kleinknecht, 2016). Videos realistically portray the demands of the teaching professions and therefore allow prospective teachers to develop situated aspects of competence: the description, interpretation, and evaluation of realistic teaching situations (Krammer, 2014; Moreno & Valdez, 2007). The case offers a 'window' into the classroom without the added pressure of having to immediately act and interact in the classroom situation (Sherin, 2007b). Learning through the analysis of cases is considered to have many advantages (Blomberg et al., 2013). Firstly, cases present a suitable method for the connection of theory and practice. Secondly, cases present the possibility to explore alternative teaching strategies and teacher agency. Thirdly, the analysis of cases allows prospective teachers to explore the richness and complexity of practice and to discuss their perspectives and experiences. Fourthly, the teacher of the course has a large degree of control over the topic and direction of the learning situation, because he or she can prepare the case in any way. Fifthly, cases increase the motivation for learning. Sixthly, instead of overwhelming learners with the full complexity of teaching, cases focus the attention on a limited amount of aspects that are considered relevant to the topic of the course (Shulman, 1992).

A case represents an example of a typical situation, "an instance of a larger class, an example of a broader category" (Shulman, 1992, p. 17). The cases can be related to a concept in two different ways: *deductively*, in which the case is presented as a means to apply or recognize a previously presented concept, or *inductively*, in which the case is presented as a means to formulate a more general principle (Goeze, Hetfleisch, & Schrader, 2013). The function of the case is different for both methods. In the deductive strategy, a case is used to focus the attention of the learners on the concept under scrutiny. In the inductive strategy, a case is used to focus the attention of the learners to take the perspective of the involved subjects (Goeze, 2010).

4.3.2 Video-cases.

In comparison to text-based cases, the format of video-recordings is assumed to provide a higher degree of authenticity, especially regarding the complexity of the teaching context of the case (Putnam & Borko, 2000). This is assumed to be especially relevant for acquiring classroom management competence that is characterized by *multidimensionality*, *simultaneity*, *immediacy*, *unpredictability,* and *history* (Doyle, 1986). Indeed, several studies have shown the effectiveness of analyzing video recordings in university-based courses (Hellermann et al., 2015; Kramer et al., 2017; Kumschick et al., 2017). Learning with videos is associated with high degrees of motivation and a feeling of immersion (Syring et al., 2015). Moreover, it was shown that courses using videos tended to lead to higher degrees of cognitive activation of learners (Kramer et al., 2017). However, this additional complexity and feeling of immersion make it more likely that novices feel overwhelmed and face high degrees of cognitive load (Syring et al., 2015). Therefore, it is generally assumed that novices, especially at the beginning of their studies, profit from courses with a higher degree of structure regarding the learning materials and learning support (Santagata & Guarino, 2011).

For the effective implementation of video-based case-analysis in teacher training courses, many different methodologies have been developed (Gaudin & Chaliès, 2015, p. 48). Blomberg et al. (Blomberg et al., 2013) summarized these methods through the formulation of several research-based heuristics. These heuristics serve as guidelines for the development of video-based courses.

Learning objectives

The formulation of appropriate learning objectives is an important first choice. Video-based courses should not be considered a panacea. To Blomberg et al. (Blomberg et al., 2013), suitable learning objectives are firstly *selective attention*, helping prospective teachers to filter relevant from irrelevant classroom interactions. Secondly, knowledge-based *reasoning:* fostering a deeper level of interpretation of the teaching aspect under scrutiny, application of theories or teaching strategies, evaluation of teacher behavior or predicting of outcomes. More specific learning objectives are firstly related to the type of video-cases and the instructional strategy with which

these video-cases are incorporated in the course.

Video-cases

Blomberg et al. (Blomberg et al., 2013) additionally point at other decisions related to the properties of the video cases and their implications for the learning objective. Firstly, courses can use videos of learners' *own* teaching or videos of *external* or unknown teachers. Analyzing one's own teaching is assumed to increase the sense of immersion and therefore, strengthen the potential of self-reflection (Gaudin & Chaliès, 2015). Research shows that analyzing one's own teaching indeed leads to higher emotional-motivational involvement, immersion, and motivation (Seidel, Stürmer, Blomberg, Kobarg, & Schwindt, 2011). Although this involvement might lead to more self-reflection, research indicates that this increased involvement might also hinder the effectiveness in other areas. It was shown that teachers analyzing unknown teaching situations were more engaged in the analysis of problematic events. An increased distance due to the depicting of unknown teaching might also increase the learning effects regarding the application of theories (Kleinknecht & Schneider, 2013). Videos of *familiar* contexts are easier to interpret because viewers can identify with the depicted teaching setting. However, teaching depicting an *unfamiliar* situation will lead to more discussion, may require additional scaffolding to make sure its content is accurately interpreted.

Secondly, courses can use videos depicting *best*, *typical* or *bad* teaching practice. Blomberg et al. (Blomberg et al., 2013) suggest that this decision depends on the learning objective and instructional strategy. When the aim is to show novel or effective ways of dealing with teaching problems, best practice video should be used. Alternatively, when the aim is to engage learners in reflection on their teaching, videos of typical practice should be used.

Thirdly, an increasing number of studies use *staged* instead of *authentic* videos (Piwowar, Barth, Ophardt, & Thiel, 2017). The majority of video-based case-analysis is done with authentic cases. Because in recent years, many video-survey studies were held, many authentic cases are available (Seidel & Thiel, 2017). Another reason is the high costs involved in the production process of staged videos (Barth, 2016). The main advantage lies in the fact that no issues concerning privacy exist (Rauin, Herrle, & Engartner, 2016). Moreover, the script can focus the

attention of the learners to specific interactions. Studies have shown that also staged videos are considered to be engaging, motivating (Piwowar et al., 2017) and can be effectively used for purposes of teacher training (Kumschick et al., 2017).

4.3.3 Instructional strategies.

The fourth and final research-based heuristic that Blomberg et al. (Blomberg et al., 2013) include, refers to the difference between using cases deductively or inductively, as described in chapter 4.3.1 (Goeze et al., 2013). When designing learning arrangement, one must choose a so-called instructional strategy: embedding video clips following either a *situated* or a *cognitive* instructional strategy (Blomberg et al., 2013). These strategies are based on two different theoretical perspectives on the mechanisms of learning (Nückles & Wittwer, 2014): on the one hand the *cognitive-constructivist* perspective which focuses on reducing cognitive load by offering highly structured instructional designs. And on the other hand, the *situated-learning* perspective that focuses on offering learning arrangements that are as realistic as possible and therefore increase the probability of transfer. The assumptions underlying these perspectives have important implications for the way videos should be used to foster the learning processes, especially regarding the degree of complexity that is assumed to be appropriate for novice learners. Many textbooks present these perspectives as opposing, mutually excluding (Krapp & Weidenmann, 2006; Rost, 2001). Describing these perspectives in their ideal type helps to understand their difference. However, in practice, the conceptual borders between the perspectives are blurry and even changed over time as both perspectives were developed further (Reinmann, Gabi & Mandl, Heinz, 2006). Moreover, it is generally assumed that both perspectives are valid and can complement each other (Nückles & Wittwer, 2014).

Cognitive instructional strategy

In cognition science, learning is understood as the successful acquisition of knowledge. This knowledge-acquirement is thought of as a process of information processing. For the description of this process, the metaphor of mechanical information-processing of a computer is used (Woolfolk, 2005). The process of learning is considered successful when the related information is successfully processing and stored as knowledge in long-term memory. The mechanism with

which information is processed is referred to as the *Information-Processing Model of Memory* (see Hasselhorn & Gold, 2013) and consist of three steps. Firstly, External stimuli are received and transformed through the senses and for the duration of the stimulus stored in *sensory registers*. When the learner focuses his or her conscious attention to these registers, the information is stored in *short-term* or *working memory.* Through conscious and systematical practice and structuring of the information stored in short-term memory, the knowledge is effectively stored within an infinite amount of *long-term memory*. For successfully saving information in long-term memory, the amount of working memory is considered the bottleneck, because of its limited capacity and short duration (Sweller, 2002).

According to constructivist psychological theory (Seidel & Krapp, 2014), the learning subject does not passively receive and store information. Instead, the subject actively selects the stimuli to process, depending on its conceived subjective significance. Learning is assumed to be the result of active construction, in the sense that new knowledge is actively connected to the knowledge that is already stored in long-term memory. Constructivist learning, therefore, involves bringing up knowledge, as well as storing new knowledge. Knowledge is stored in and brought up from long-term memory in the form of *schemata*. Schemata refer to the building blocks of knowledge and represent abstract mental representations through which meaning is assigned to stimuli. The repeated analysis of an object leads to the discovery of regularities and core qualities. Through practice, learners can refine the schemata and improve their ability to discern important from irrelevant qualities of that object. Schemata that refer to rule-based sequenced action-patterns are referred to as *scripts*. The ability to solve complex problems is assumed to be based on “tens of thousands of domain-specific schemas” (Sweller, 2002, p. 1503). When the process of alignment of stored schemata with new problems or cases is repeated, this process can become automated. Then, the process of alignment can occur unconsciously so that working memory becomes available for the processing of other alignment processes. However, when insufficient schemata are available, or these schemata are not sufficiently automated, the processing of information alignment claims a large part of available working memory.

Cognitive load theory can be considered a specification of the cognitive-constructivist perspective (Sweller, 1988). It assumes that learning consists of the alignment of schemata in

short-term memory but adds the degree of cognitive load as a decisive factor for the successful storage of knowledge in long-term memory (Sweller, 1988). In cognitive load theory, it is assumed that short-term memory is the bottleneck within the learning process because it is easily strained. This strain is referred to as cognitive load. Three kinds of cognitive load are discerned (Sweller, Van Merriënboer & Paas, 1998): *Intrinsic* cognitive load relates to the complexity of the knowledge that is being acquired. *Extraneous* cognitive load is a result of the structure of the instructional design through which knowledge is conveyed. Well-structured designs can significantly reduce the cognitive load with learners and therefore increase the probability that the information contained in stimuli can be stored. *Germane* cognitive load is the load resulting from the constructing of mental schemata through consciously processing of information.

According to this logic, learning can only occur when the cognitive load resulting from the learning process does not exceed the available capacity of the working memory. Moreover, it follows that learning depends on three factors: the learner, the object of learning, and the instructional design of the learning arrangement (Sweller et al., 1998). These three factors are explained in more detail below.

Firstly, effective learning depends on the learners. Learners differ in the degree to which their short-term memory becomes overloaded. Learners, for example, have different degrees of automated schemata related to the subject of learning are available. When existing knowledge in the form of automated schemata is available, it reduces the amount of germane cognitive load.

Secondly, learning depends on the complexity of the object of learning. Information within objects of learning is structured in so-called *elements*. Depending on the complexity of the object of learning, elements differ in their degree of interactivity. Elements with a low degree of interactivity, like for example, the vocabulary of a foreign language can be learned serially because the effective learning of one word is not based on the learning of another word. However, elements with a high degree of interactivity, like for example the grammatical properties of a foreign language, can only be learned simultaneously. The learning of elements with high interactivity can only occur simultaneously and therefore involves a high degree of intrinsic cognitive load.

Thirdly, learning depends on the instructional design of the course. One of the main responsibilities of a teacher consists of designing courses in such a way that extraneous cognitive load is minimized so that the probability a learner can store the new information in long-term memory. Instructional designs should, therefore, firstly encourage the construction of schemata and automation. Secondly, courses should be systematically and meticulously planned in so-called *instructional designs* (Reinmann, Gabi & Mandl, Heinz, 2006). The focus of these instructional designs lies with the optimization of instruction so that extraneous cognitive load is minimized so that the probability that complex learning can be successful. Some of the guidelines following this logic involve simple-to-complex sequencing, just-in-time information, the availability of procedural information, and the avoiding irrelevant information (Van Merriënboer, Krischner, & Kester, 2003).

Situated instructional strategy

From a situated perspective, learning is also considered to be the result of the process of knowledge acquisition. However, instead of the factor cognitive load, the situated perspective puts a focus on the factor social context (Seidel & Krapp, 2014). Within situated learning, knowledge is not seen as an abstract entity that is independent of its social context. Instead, knowledge is principally connected to and embedded in the social context in which this knowledge is acquired. The transfer of knowledge between different social contexts is therefore difficult because acquired knowledge is only valid within that particular context (Gruber, Law, Mandl, & Renkl, 1996).

When the social practice in which knowledge is acquired is different from the social practice in which the knowledge is intended to be applied, so-called *inert knowledge* is created (Klauer, 2001). Inert knowledge refers to knowledge that is successfully acquired, but that cannot be accessed in a different context. To situational theorists, this applies to the majority of the knowledge acquired in school. Knowledge, acquired in schools, can only be reproduced in that same school setting, like for example in a test. According to situated learning theory, it is wrong to assume that a transfer of book-knowledge to practice takes place more-or-less automatically (Klauer, 2001).

The situated perspective on learning implies the dismissal of the distinction between the personal inner cognitive processes and the outer, social world (Greeno, 1998). This means that learning is seen as a social practice of participating in so-called *communities of practice* (Lave & Wenger, 1991). Taking a social-interaction perspective, the meaning is assumed to be the product of the sharing of goals, values, and rules of interaction within a specific social context (Putnam & Borko Hilda, 2000). This explains why the transfer of knowledge is not always successful. When for learning, knowledge is abstracted from the social a physical context in which it is aimed to be applied, this knowledge loses its meaning and becomes inert. Hence, the proponents of the situated perspective aim at placing the acquirement of knowledge as 'closely' as possible to the context of future application. Therefore, effective learning should be based on the practicing within authentic activities (Putnam & Borko Hilda, 2000, p. 4). It should be focused on the solving of authentic problems, with authentic methods. The form of the exercises should be close to real-life. Moreover, learning should take place within complex social systems, as a reduction of complexity would represent a further detachment of the social setting. (Reinmann, Gabi & Mandl, Heinz, 2006).

Comparison

The situated and cognitive perspectives differ strongly regarding their central assumptions of effective learning. Indeed, several studies aimed at refuting the central claims of the other perspectives (Anderson, Reder, & Herbert, 1996; Kirschner, Sweller, & Clark, 2006). However, there are also studies that seek to bridge the gap between the two perspectives. One academic initiative brought together proponents of both theoretical perspectives to jointly formulate four points on which both perspectives do *not* differ (Anderson, Greeno, Reder, & Simon, 2000). Firstly, both perspectives should not be mistaken to have different units of analysis: both individual and social aspects of learning are important in education, and both perspectives do not imply a focus or preference on either aspect or learning activity. Secondly, learning can be accomplished on a general abstract level, but sometimes it cannot. A transfer of knowledge from one situation to the other is possible. However, this transfer cannot be assumed to take place automatically. Thirdly, both perspectives view the processes of learning differently, but both perspectives make contributions in educational sciences and should be pursued in future

research. (Klauer, 2001). Fred Korthagen (Korthagen, 2010) adds that both perspectives have different functions, stating that "situated learning theory tries to explain the role of embodied social learning, while cognitive theory aims at describing the characteristics of knowledge and knowledge development per se" (Korthagen, 2010, p. 99). It is therefore concluded that, principally, both instructional strategies can be used for the design of courses in which video-clips are analyzed. Indeed, several studies have compared the differences in learning effects of both instructional strategies.

4.3.4 Effects of video-based courses in university.

Many studies have shown the positive effects of analyzing videos in courses for prospective teachers (Gaudin & Chaliès, 2015). In this section, we firstly discuss the effects of video-based courses on specific competence aspects. Then, the differential effects of the instructional strategies (Steffensky & Kleinknecht, 2016) is discussed.

Situated competence aspects

The large body of empirical studies show the effects of video-based case-analysis on situated competence aspects of prospective teachers (Steffensky & Kleinknecht, 2016) like *professional vision* (Gold, Förster, & Holodynski, 2013; Sherin & Van Es, 2005; Star & Strickland, 2008), the *ability to analyze* (Santagata & Guarino, 2011; Santagata & Yeh, 2014), and *situation-specific skills* (Stahnke et al., 2016). Studies reviewing the effects of video-based courses (Steffensky & Kleinknecht, 2016) show that firstly, prospective teachers improve their *selective attention* or the ability to filter out relevant information. Before participation in video-based courses, prospective teachers tend to focus on superficial aspects that are less important for the teaching situation. After course-participation, prospective teachers can better filter out irrelevant information and better focus on the relevant aspects of students' thinking and identification of critical classroom interaction. Secondly, prospective teachers change their level of *knowledge-based reasoning.* Before a course, prospective teachers mainly stick to describing the situation depicted in the case. After a course, participants can additionally interpret the meaning of the case regarding relevant theory and can evaluate the effectiveness of teacher's interventions or the predicting of the outcome. Furthermore, empirical studies also indicate that participation in

video-based courses changed the teaching practices of prospective teachers in their internships. Exploratory research on prospective elementary mathematics shows that prospective teachers who participated in video-based courses provided their students with more opportunities to explain their thinking processes (Santagata & Yeh, 2014; Sun & Van Es, 2015).

Instructional design

Building on the assumption that both a *situated* and a *cognitive* instructional strategy are potentially valid for video-based case-analysis (Blomberg et al., 2013), several studies have systematically compared the effects of the two instructional strategies. In the following section, the results of four studies are summarized. The number of available empirical studies systematically comparing the instructional designs is currently limited and inconsistent. Indeed, the lack of empirical studies is mentioned in each of the following studies. Additionally, the four studies discussed below operationalize the theoretical differences between both strategies differently. However, because the theoretical assumptions are similar, the discussed studies are assumed to reflect a similar systematic difference in course design.

A first study focused on the effects that instructional strategy has on the emotional and cognitive processes of prospective teachers (Syring et al., 2015). A total of n = 641 prospective teachers in the second semester of their secondary teacher training track were non-randomly assigned to a course using a *problem-based* situated instructional strategy or a *direct-instruction* cognitive strategy. In the course using a problem-based strategy, prospective teachers worked independently in groups on the text- and video-based cases, with the teacher having a more remote and moderating role. In the course based on direct-instruction, the teacher demonstrated a step-by-step case-analysis after which prospective teachers worked individually on tasks using the explained methodology. The study showed that contrary to what was expected from theory, *extraneous* cognitive load was found to be higher in the course based on direct-instruction. The authors explain this difference by stating that the participants had to concentrate more on following the steps as presented by the teacher. In both experimental groups, intrinsic cognitive load decreases over time, suggesting that the participants get accustomed to the case-analysis method. Regarding general motivation, no significant differences could be found

between the groups. Regarding the sense of *immersion*, it could be shown that the problem-based courses evoked a significantly stronger sense of immersion with a small effect size. The authors conclude that both instructional strategies are suitable for the use of video-based case-analysis for prospective teachers. Because the result shows that problem-based courses tend to show slightly lower values of cognitive load and slightly higher values of immersion than direct-instruction, the authors advise teacher training courses to consider using a problem-based situated instructional strategy. The authors also note that these results are limited to prospective teachers at the beginning of their teacher training program and that different results are expected for prospective teachers in a more advanced state of their training program.

A second study conceptualizes the instructional strategies as either Rule-Example or Example-Rule (Seidel et al., 2013). Within learning arrangement using an example-rule instructional design, videos are used as a problem anchor from which general rules and strategies can be deduced. With a rule-example instructional strategy, videos are used as examples to illustrate and consolidate previously conveyed rules and strategies. The sample consisted of n = 54 prospective teachers in their second year of training who were randomly assigned to one of the two courses. Both courses consisted of eight sessions, each session lasting 90 minutes and discussed three components of teaching quality: *goal setting*, a*ctivation of student thinking*, and *providing a supportive learning climate*. Results show, firstly, that regarding declarative knowledge the rule-example group performed significantly better than the example-rule group with a large effect-size (p < .05; d = 1.86). Secondly, regarding professional vision, the rule-example group performed significantly better than the example-rule group with a medium effect-size (p < .05; d = 0.55). Thirdly, the example-rule group performed significantly better regarding lesson planning with a large effect-size (p < .05; d = 0.91). The researchers conclude that the learning goal should determine which instructional strategy should be used. For the acquisition of factual knowledge, a rule-example strategy should be used. However, the example-rule strategy should be used when the aim is to foster the application of knowledge.

A third study conceptualized the systematic difference between courses in line with the situated and cognitive view on learning (Blomberg et al., 2014). A sample of n = 28 prospective teachers in their second years of their bachelor's program was randomly assigned to one of the two

courses. Both courses introduced three components of instructional quality in school teaching: *goal setting*, a*ctivation of student thinking*, and *providing a supportive learning climate.* The ability to reflect on their learning was chosen as a dependent variable. For the assessment of the ability to reflect, course-participants were asked to write down their reflections concerning that day's session in a learning journal. They were asked to write learning journals and reflect on what they found difficult to understand and what they might do to improve their understanding. Data-analysis showed that the group participating in the course following the cognitive challenge initially showed more expert level reflection on the level of *integration*. However, this could not be sustained over time and gave way to more *description* in the final learning journals. Alternatively, the group in the course following a situated strategy could sustain the number of reflections on the level of *description,* but could also increase their amount of reflections on the level of *integration*. The number of reflections on the level of *evaluation* went down for both groups. The authors conclude that prospective teachers initially benefit from higher degrees of course structuring when analyzing videos, but that when fostering the ability to reflect in the long run, a situated strategy might be more effective.

A fourth study systematically compared the instructional strategy on classroom management competence (Kumschick et al., 2017). A sample of *n* = 237 prospective teachers in their master's program was assigned to two treatments: *n* = 113 prospective teachers were assigned to a classical problem-based course (PBL); and *n* = 124 prospective teachers were assigned to a so-called 'instructed problem-based' course (I-PBL). The courses lasted for four sessions of 90 minutes. Both courses only differed in the first two sessions. In the first session, the PBL group watched a staged bad-practice teaching case related to classroom management. In the second session, theoretical knowledge was gained by autonomously and cooperatively working on a reader. By contrast, the I-PBL group received direct instruction in the first session and during the second session, in which theoretical knowledge and classroom management tools were presented. The third and fourth sessions of both groups were equivalent, in which both groups worked on analyzing the bad-practice and a best-practice classroom management case. Results showed a strong significant main effect on perceived self-efficacy, but no significant interaction effects could be found, indicating that both groups showed a similar increase in their self-efficacy.

Results of the comparison of declarative knowledge in posttest showed that the I-PBL group scored significantly higher than the PBL group with a medium-sized effect. The authors conclude that the results show that a problem-based approach to learning with video-cases is effective for use in programs for prospective teachers. However, the results of the knowledge-test imply that its effectiveness on declarative knowledge can be strengthened when this problem-based approach is accompanied by direct instruction.

Although referring to the same theories of the situated and cognitive approach, these four studies have operationalized these strategies differently. The first study (Syring et al., 2015) compared direct instruction to problem-based learning. The second study (Seidel et al., 2013) compared rule-example to example-rule, the third study (Blomberg et al., 2014) compared a cognitive to a situated perspective, and the fourth study (Kumschick et al., 2017) compared an instructed problem-based approach to a problem-based approach. Moreover, these studies differ in other relevant variables, like for example social form (group work versus individual work), amount of structuring by the teacher (step-by-step guidance vs independent exercise by the learners), the role of available information (supportive information vs indirect instruction) and learning activities (application vs deriving of rules or teaching strategies). Therefore, it becomes difficult to attribute the effects only to the difference in instructional strategy. Nevertheless, these studies indicate the following. Firstly, all studies showed a positive learning effect for both instructional designs. Secondly, for perceived self-efficacy of classroom management, no significant differences between the two strategies could be found. (Kumschick et al., 2017). Thirdly, cognitive competence aspects of competence are more strongly fostered with a cognitive strategy (Kumschick et al., 2017; Seidel et al., 2013), whereas a situated strategy was more effective on situated competence-aspects like the formulation of lesson-planning (Seidel et al., 2013). Regarding the ability to reflect, it was shown that a cognitive strategy produced more expert-like reflections at the beginning of the course, whereas a situated strategy lead to more expert-like reflections at the end of the course (Seidel et al., 2013). Fourthly, confirming the theoretical assumption of the situated strategy, courses based on problem-based learning lead to a stronger sense of immersion with learners (Syring et al., 2015). Fifthly, contrary to the cognitive load theory, a course based on direct instruction did not lead to lower degrees

of external cognitive load (Syring et al., 2015).

4.4 Conclusion

Chapter four aims at providing an overview of different methods with which prospective teacher acquire classroom management competence. The main advice from research on classroom management for university-based learning arrangements is to focus on including the teaching practice, either through internships (Larrivee, 2006) or the analysis of video-based case-analysis (Jones, 2006).

Internships have a broad range of effects on prospective teachers' competence. However, these effects are heterogeneous and tend to differ strongly between prospective teachers (Hascher, 2006, 2012a). On average, prospective teachers tend to increase their declarative general pedagogical knowledge, their self-efficacy, their situation-specific skills, and gain a more realistic perspective of school life and the role of the teacher.

A large body of research shows that video-based courses are effective for the conveying of situation-specific skills (Gaudin & Chaliès, 2015). These effects were shown both for the domain of general pedagogical knowledge (Steffensky & Kleinknecht, 2016) and the specific domain of classroom management (Hellermann et al., 2015; Kramer et al., 2017). Because the analysis of videos is associated with high degrees of complexity (Santagata & Guarino, 2011), it is expected that case-analysis of video recordings pose a challenge for prospective teachers. Following cognitive load theory, it is assumed that learning arrangements should be highly structured, providing learners with step-by-step guidance and relevant theories. However, situated learning theory assumes a strong reduction of complexity creates distance between the social context in which the learning takes place from the social context in which the achieved knowledge should be applied. Therefore, video-cases' complexity should not be reduced as this would reduce the degree of authenticity (Greeno, 1998). Instead, the meaning of knowledge should be constructed socially and based on a permanent reflection on previous experiences and subjective theories (Korthagen, 2010).

Regarding the analysis of video-cases, empirical research suggests that the choice of instructional

strategy should be dependent on the learning objective (Blomberg et al., 2013; Blomberg et al., 2014; Kumschick et al., 2017; Seidel et al., 2013; Syring et al., 2015). These studies view prospective teachers as a relatively homogenous group. However, research on the effectiveness of internships gives reason to believe that there are relevant differences between prospective teachers with and without practical experience. Therefore, it is argued in the current study, that the choice for either instructive strategy should be dependent on the experience, knowledge, situation-specific skills, and beliefs of learners. The research on the effects of in internships indicates that during internships, prospective teachers acquire situation-specific skills, increase their general pedagogical knowledge, and develop a better understanding about school life and the role and daily work of teachers. It is assumed that these effects play an important role for the choice for one of the two instructional strategies. To be able to recognize that a case is an example of a more general concept and successfully make use of a deductive approach, as proposed by the situated instructional design, a learner must be aware of similar cases. This awareness only exists, or has adequate validness, when a prospective teacher has participated in field experiences like for example an internship.

Also, it is assumed in the current study that the degree to which learners are exposed to cognitive load depends on the amount of familiarity with similar schemata and scripts. Prospective teachers who have minimal experience within school will be less acquainted with the classroom settings from a teacher's perspective and therefore more easily suffer from cognitive load. In the current study, therefore, it is assumed that the choice for a situated deductive or cognitive inductive instructional strategy should depend on the availability of practical experience of the learners. A situated strategy is more effective for prospective teachers who have already gained experience in an internship for two reasons. Firstly, their short-term-memory will be less strained as they can more easily align new information with available schemata and scripts. Secondly, they have beliefs and hold subjective theories that are more in line with realistic teaching and can, therefore, form the basis of the reflection of video-cases. Alternatively, a cognitive strategy is expected to be more effective for prospective teachers without practical experience and no participation in internships. Without this experience, cognitive load of learners will be much higher. Therefore, they will profit more from strongly structured courses.

5. Research Question

This chapter firstly summarizes the main points of the preceding chapters. Secondly it lists the respective research desiderata. Thirdly, the research questions of the current study are listed. The chapter is structured according to the three aims of the study (see chapter 1): to contribute to the discussion on how prospective teachers' classroom management competence should be modeled (5.1), to contribute to the discussion on the way in which prospective teachers' situated competence should be measured (5.2), and to contribute to the discussion on how to effectively convey classroom management competence of prospective teachers through video-based courses (5.3). The chapter is concluded by stating the general research question and subquestions for each of the three above-listed aims (5.4).

5.1 How to Model Classroom Management Competence

The teaching aspect of classroom management was defined as "the actions teachers take to create an environment that supports and facilitates both academic and social-emotional learning" (Evertson & Weinstein, 2006b, p. 4). As such, classroom management is considered to consist of the in-situ teachers' behavior, independent of the teaching subject, and supportive of student learning, rather than representing a learning objective itself (see chapter 1). The ability of a prospective teacher to apply effective strategies for the prevention of classroom disturbances is of central importance for student achievement and teacher well-being. Research on classroom management identifies three foci: the management of group interaction, the increasing of the effective use of available learning time, and managing lesson disturbances (see chapter 2.3).

Classroom management competence needs to be modeled for valid measurement and to effectively strengthen prospective teachers' classroom management competence within university-based courses (see chapter 3.1). A dispositional approach to competence that focuses on cognitive competence-aspects is expected to be insufficient for the modeling of classroom management competence, because declarative knowledge plays a relatively minor role within classroom management competence (see chapter 3.2). However, also a behavioral approach is

expected to be insufficient. This approach does not provide a starting point regarding the methods with which prospective teachers with minimal teaching experience can strengthen classroom management competence in university-based courses. Instead, a situated approach to teacher's competence, the PID-model, is used to model prospective teachers' classroom management competence (see chapter 3.4). The PID-model identifies perception, interpretation, and decision-making, so-called *situation-specific skills,* in the middle of the two ends of the competence-continuum. On the one hand, these situation-specific skills are assumed to build on dispositional competence-aspects like knowledge, beliefs, and motivational dispositions. On the other hand, situation-specific skills are assumed to be predictive for classroom management performance in real life teaching situations.

Two research initiatives, CME and VIU, have applied a situated approach to the modeling and measuring of classroom management competence (see chapter 3.5). These studies have shown that classroom management competence can be reliably measured with prospective teachers through the use of video tests. However, in both studies, the dimensional structure that discerns between the situation-specific skills of description and interpretation could not be empirically confirmed. Therefore, the question remains open whether these situation-specific skills represent independent aspects of classroom management competence of prospective teachers.

Several studies indicate that performance on video tests could be dependent on the personal trait of attentiveness, the degree to which a person can successfully process visual information (see chapter 3.4). This would mean that the variance in the performance on video tests is in part ascribed to a personal trait instead of an aspect of a person's competence. No studies were found that investigate to what degree a score on a video test correlates to a person's attentiveness.

Although not an aspect of competence that can be objectively measured, it was shown that teachers' self-efficacy in classroom management is of high relevance for conveying classroom management competence in university (see chapter 3.7). On the one hand, several studies showed that self-efficacy is strongly related to starting teachers' emotional exhaustion and burnout. On the other hand, one study shows that low rates of emotional exhaustion for beginning teachers are a prerequisite for the transfer of knowledge from their university

programs into the teaching practice. Moreover, several studies have shown that both internships (see chapter 4.2), as well as video-based courses, can increase prospective teachers' degree of self-efficacy (see chapter 4.3). This implies that both can be effective for the reduction of stress in the beginning stages of teaching and can, therefore, reduce the reality shock, underlining the importance of self-efficacy for the current study.

5.2 How to Measure Classroom Management Competence

Recently, many projects were initiated in which situation-specific skills were measured using a video test (see chapters 3.3, 3.4, and 3.5). However, it was also shown that the systematic comparison between these different initiatives is problematic as researchers tend to leave out the specifics of their test-design (see chapter 3.8). It is the aim of the current study to provide as much transparency as possible regarding the process of test-design.

5.3 How to Convey Classroom Management Competence in University

Universities typically use video-based courses and internships to convey classroom management competence for prospective teachers (see chapter 4), as they are assumed to be appropriate for a teaching-aspect that is strongly related to the teaching context (see chapter 4.1). Several studies have investigated the effects of participation in internships (See chapter 4.2). Regarding non-cognitive competence-aspects, participation in an internship tends to strengthen self-efficacy, change subjective theories, and make prospective teachers develop a more realistic view of teaching and school life. Moreover, participation in internships increases prospective teachers' general pedagogical knowledge, positively influence situated competence aspects within the domain of general pedagogical knowledge, and strengthen prospective teachers' perceived self-efficacy. However, no studies were found that could confirm whether participation in an internship has similar effects on aspects of classroom management competence.

Empirical studies have repeatedly confirmed the effectiveness of using videos in teacher training programs (see chapter 4.3). Also, for the conveying of classroom management competence, the use of video-based case-analysis proved to be an effective method (see chapter 4.3.3). When designing a video-based course, one of the choices is the instructional strategy: the way video

clips are didactically embedded in the course. The situated strategy uses videos as a way to activate previous experiences based on which regularities, strategies, and theory can be deduced. This strategy is based on the theory of situated learning (see chapter 4.3.3). Situated learning theory assumes that learning should be located as closely as possible to the practice in which the acquired competence should be applied. Disconnecting the learning setting from the intended application setting, for example, through the strong methodological structuring, increases the risk of failed transfer between the social contexts. Instead, learners should be immersed in the initial complexity of the teaching setting, as it represents the only valid social context in which the acquired competence is later to be applied. Alternatively, within a cognitive instructional strategy, video clips are used to apply previously acquired theoretical knowledge through a pre-defined strongly structured instructional strategy. This cognitive strategy is based on cognitive load theory (see chapter 4.3.3), in which it is assumed that learning can only take place when levels of cognitive load, typically associated with learning with complex materials, like video clips, are not too high. As a consequence, to decrease levels of cognitive load, video-based courses should initially include high levels of structure and should aim at automation.

Several studies show the effectiveness of (a) the instructional strategy of video-based courses and (b) the influence of teaching experience through internships on the development of prospective teachers' competence (see chapter 4). However, a research desideratum exists for the systematic investigation and possibility of interaction effects of both factors. In the current study, it is hypothesized that the choice for the use of either a situated or a cognitive instructional strategy should be dependent on the availability of teaching experience (see chapter 4.4). For the situated strategy to be effective, a realistic perspective on teaching and school life, as well as a realistic notion of the teaching experience, must be present. Therefore, for prospective teachers in an advanced stage of their program, including the participation in an internship, a situated approach is assumed to be more effective. Alternatively, for prospective teachers that are in the first stage of their program, it is assumed that the levels of cognitive load are much higher. Because these learners have not yet acquired teaching experience in the practical semester and are therefore not sufficiently acquainted with teaching, video-based courses following a cognitive instructional strategy are assumed to be more effective. Only a handful of

empirical studies were found that have systematically assessed the differentiated influence of instructional strategies (see chapter 4.3.4). None of these studies have investigated the impact of available teaching experience gained through participation in an internship.

5.4 Research Questions

The primary objective of the current study is to investigate how video-based courses can contribute to prospective teachers' classroom management competence.

5.4.1 Main research question.

Therefore, the study focuses on investigating the influence of the availability of teaching experiences through participation in a practical semester and the way that the instructional approach of the video-based course can build on these differences in teaching experience. Following this objective and the research desideratum, as presented in the preceding sections, the main research question is formulated as follows:

How can classroom management competency be effectively conveyed in video-based courses of teacher training programs regarding the differences in practical experiences of prospective teachers?

and the following hypotheses were derived:

- Prospective teachers with experience in an internship benefit more from a situated instructional strategy.
- Prospective teachers without experience in an internship benefit more from a cognitive instructional strategy.

5.4.2 Sub-questions.

Following the desiderata in the preceding sections, the following sub-questions were formulated.

<u>Theory: How to model classroom management competence</u>

1. Do the data confirm the dimensional structure of the PID-model in which *perception* and *interpretation* are two different aspects of situated competence?

2. To what degree does the trait of attentiveness play a role in the performance of video tests?
3. To what degree does self-efficacy change as a result of the participation in the courses on video analysis?

Methodology: How to measure classroom management competence

4. How can classroom management competence be validly measured with prospective teachers with a newly developed video test?

Praxis: How to convey classroom management competence

5. How can university-based courses be designed according to the guidelines of situated and cognitive instructional strategies?
6. To what degree are practical teaching experiences, as gained through an internship, relevant for the ability to analyze video clips?

6. Methodology

The first part of this chapter describes the process of developing a new video test that measures Classroom Management Competence (CMC). It was decided to develop this new test, because both available video tests, VIU and CME (see chapter 3.6), were expected to be inappropriate to test the main hypotheses. In CME, the construct of classroom management was only operationalized through the classroom situations depicted in the video clips (see chapter 3.6.3). For item-formulation, only the three cognitive challenges of *Accuracy of Perception* (AP), *Holistic Perception* (HP), and *Justification of Action* (JA) were used (König & Lebens, 2012). No further specifications about dimensions or strategies of classroom management were used for item-formulation. As a consequence, CME represents a relatively general understanding of classroom management. For the evaluation of video courses, it is advised that the content of the course reflects the structure of the evaluation instrument (Blomberg et al., 2013). As CME does not sufficiently specify its understanding of classroom management, this test does not allow this reflection decreasing the possibility significant effects can be found.

The Items in VIU operationalize the classroom management strategies of *monitoring of student behavior*, *managing momentum*, and the *establishment of rules and routines*. However, the video clips exclusively depict classroom situations within primary education. Because the test-participants of the current study took part in the track for secondary education, it was expected that the videos violate ecological validity. For these reasons, it was decided to develop a new video test, of which the video clips, the operationalization of classroom management, and item-difficulty were purposefully focused on the evaluation of the courses and the specific population. Chapter 6.1 discusses the relevant test-theory and the different steps of the development of a video test. Chapter 6.2 presents and discusses the development process of the new video test, called CMC, and discusses the results of a pilot study.

The second part of this chapter describes the research design. Following the research question formulated in chapter 5, the research design focuses on the systematical investigation of the differentiated effect on classroom management competence of the following two factors. The first factor, *cohort,* consists of two groups: a group of prospective teachers that have not yet

participated in the practical semester and a group of prospective teachers that have already participated in the practical semester. The second factor, *course-type* also consists of two groups: a group participating in a course developed according to the principles of a cognitive strategy and a group participating in a course developed according to the principles of a situated strategy. Combining these two factors results in four groups. The construct of classroom management competence is quantitatively measured in pretest and posttest using the video test called CMC. Hypotheses about the effects are derived from theory, and the sample can be considered to be non-probabilistic self-selected. Summing up, the design of the research project can be characterized as a (a) *quantitative*, (b) *explanatory*, (c) *quasi-experimental* (d) *field study* with a (e) *repeated-measurement,* (f) and a *2x2, factorial* design (Döring & Bortz, 2016b, p. 183). Figure 5 presents a graphical overview of the research design.

<table>
<tr><th>Cohort</th><th>Oct. ‘15 – Feb. ‘16</th><th colspan="3">Apr. ‘16 – Aug. ‘16</th><th>Oct. ‘16 – Feb. ‘17</th></tr>
<tr><td rowspan="2">1.</td><td>Third Term</td><td colspan="3">Fourth Term</td><td>Fifth Term</td></tr>
<tr><td>Practical Semester</td><td>PRETEST</td><td>Situated Strategy
Cognitive Strategy</td><td>POSTTEST</td><td></td></tr>
<tr><td colspan="6"></td></tr>
<tr><td rowspan="2">2.</td><td>First Term</td><td colspan="3">Second Term</td><td>Third Term</td></tr>
<tr><td></td><td>PRETEST</td><td>Situated Strategy
Cognitive Strategy</td><td>POSTTEST</td><td>Practical Semester</td></tr>
</table>

Figure 5. **Research Design of the Current Study**

Chapter 6.3 describes the treatments, both the content of the courses and the practical semester. Chapter 6.4 describes all research instruments and their respective functions that were administered additionally to CMC. Further discussion of the research design consists of a description of the sample (6.5), the way guidelines for research ethics were implemented (6.6), the formulation of research hypotheses (6.7) and the methods for data analysis that were applied (6.8).

6.1. Test-Theory and Test-Development

As shown in the last section, the pilot-assessment of CME showed several inadequacies for the application in the main study. Therefore, it was decided to develop a new video test that is based on the theoretical framework, test- and item-design and coding-principles as the CME-instrument. Because of its relatively easy adaptation and the limited sample size of the main study, the development of the new video test is based on classical test theory (Bühner, 2011; Moosbrugger, 2012a; Rost, 2004). In the following sections, the main theoretical assumptions underlying classical test theory (6.1), the quality criteria with which one can evaluate and compare test-results (6.2), and the different stages of developing a psychometric test according to classical test theory (6.3) are described.

6.1.1 Classical test theory.

Classical test theory focuses on the way values of observed scores can be interpreted, and on how their true scores can be approximated (Bühner, 2011). Textbooks summarize the logic of classical test theory by listing its axioms, which can only be assumed, but not empirically verified. There are several ways these axioms are structured (Döring & Bortz, 2016b). This text firstly identifies four axioms and two additional assumptions. These are firstly described and then summarized with their respective formula (Moosbrugger, 2012a).

To classical test theory, an observed score of a test-participant *m* on a specific item *i* consists of two components: a true score and an error score. The true score t_{mi} represents the exact value of a person's trait and is considered given but unknown. The observed score x_{mi} represents the score of a test-participant, as assessed by the item. The error score e_{mi} represents the collection of internal and external factors that make observed scores vary. By definition, a true score can be approximated by determining the degree to which a specific measurement is influenced by the error score.

$$x_{mi} = t_{mi} + e_{mi}$$

A true score is approximated by holding an infinite amount of independent measurements. The value with the highest probability is referred to as the expected score $E(x_{mi})$ which is the closest

approximation of the true score.

$$E(x_{mi}) = t_{mi}$$

The error score is dependent on internal factors and external factors. Internal factors are related to the test-participant like for example his or her degree of concentration or motivation. External factors are related to the circumstances during the moment of test-administration. The direction and degree of the error-score are random. This causes the true value to be sometimes overestimated and sometimes underestimated and distinguishes unsystematic error from bias, which represents a systematical over-, or underestimation from the true value. This means that when multiple measurements are made, the average expected error variable is 0.

$$E(e_{mi}) = 0$$

The measurement of the error score is not correlated to the true score. This occurs for example when an item systematically overestimates higher true scores and underestimates lower true scores, which would also represent bias.

$$Corr(t_{mi}, e_{mi}) = 0$$

Moreover, when a test-participant processes two random items *i* and *j* repeatedly, the error between these two items does not correlate. This axiom is called *local independence*.

$$Corr(e_{mi}, e_{mj}) = 0$$

Lastly, the error between two different test-participants *m* and *v* that repeatedly process an item *i* does not correlate.

$$Corr(e_{mi}, e_{vi}) = 0$$

It follows from the assumptions of these axioms that a test-score consisting of the added scores of multiple items represents a good estimate of the true score. In comparison to item-response theory, classical test theory has four disadvantages (Moosbrugger, 2012a, p. 115). Firstly, because both error score and true score are not directly measurable, the first assumption is not falsifiable. Secondly, it is not possible to check the condition that the test-values are on the interval level of measurement. Thirdly, it is not possible to test the assumption of local

independence. Fourthly, the parameters related to classical test theory (for example those of item-analysis) are dependent on the sample. Hence, no statements on the generalizability of results to the population can be made.

6.1.2 Criteria for evaluation of a test's quality.

Objectivity, *reliability*, *validity,* and *scalability* represent the four main criteria for test quality (Bühner, 2011; Döring & Bortz, 2016b; Moosbrugger, 2012a).

Objectivity

Objectivity refers to the degree to which the results of a test are independent of both the test situation and the researcher that administers the test. There are three kinds of objectivity. Objectivity of the test-administration (a) refers to the degree to which test-scores are stable in different test situations. Objectivity of test-administration is typically increased by standardizing the duration, materials, and instruction of the test. Objectivity of scoring (b) refers to the degree to which raters use a standardized system or coding rubric that is equally interpreted and applied by each of the raters. Objectivity of interpretation (c) expresses the degree to which the interpretation of the results is independent of the test administrator.

Reliability

Reliability refers to the degree of precision to which a test or test-item measures a specific construct. In other words, reliability is high when the error score, included within the observed score, is low. There are typically four methods to calculate an instrument's reliability. Test-retest reliability (a) refers to the degree to which a score is stable over time. A test is considered reliable when the scores of two test-administration to the same sample show a high correlation. Therefore, the reliability coefficient consists of the correlation coefficient between the two measurements. This method, however, is hard to apply in practice, because a sample must be tested twice. Moreover, when the interval between the two tests is too short, learning effects may occur. Alternate-form (b) reliability refers to the method of applying two equivalent tests. This method assumes that two tests are available which can be considered equivalent. Split-half (c) reliability refers to the method with which a test is divided into two equal parts. The reliability coefficient is then calculated by correlating the test scores of both test halves under

consideration of the length of the test expressed in the number of items. Although only one test and only one assessment suffice for the calculation of this reliability coefficient, the method assumes the identification of two equivalent halves. Finally, internal consistency (d) is a similar method as the split-half method, only now each item is considered to be its own test-part. Internal consistency expresses the mean of all possible split-half correlation coefficients, or in other words, the strength of the correlation between all item-scores. Internal consistency is expressed through the Cronbach's Alpha (α) coefficient. The Alpha-coefficient is a common statistic to report test reliability and is therefore useful to compare reliability between different tests.

Validity

Validity refers to the degree to which a test measures the construct that it was intended to measure. There are three kinds of validity: Content validity (a) refers to the degree to which the content of a test, consisting of the collection of its items, sufficiently and representatively measures the aimed construct. From tests with a high level of content validity, it is assumed that the items form a representative sample from the universe of valid items. However, content validity cannot be calculated or measured. Alternatively, content validity is ensured by logical and professional considerations, mainly through the questioning of experts. Strictly speaking, only content validity fully represents the definition, presented above. However, because content validity cannot be measured, additional types of validity are needed. Criterion validity (b) refers to the degree to which a test-result correlates to the value of an external criterion: a test score of a trait that lies outside of the test situation. Depending on the time of measurement of this criterion, one differentiates between *retrospective*, *concurrent*, or *prognostic* validity. Construct validity (c) refers to the degree to which hypotheses based on the theoretical understanding of the construct of the test and related constructs are confirmed. Three kinds of construct validity can be identified. A test shows *convergent* validity when it correlates highly with a similar construct. A test shows *discriminant* validity when it shows low or negative correlations with more distant constructs. A test shows *factorial* validity when the items of a more-dimensional test can be empirically grouped into its theoretical factors.

Reliability and validity are strongly connected to each other. Firstly, because logically, the value

for reliability represents the maximum value for validity. Secondly, the reliability-validity dilemma assumes that higher values of reliability can go at the expense of validity (Rost, 2004). The dilemma has its foundation in the measurement of complex constructs. Tests for homogeneous constructs typically show very reliable measurements. However, these tests can only validly measure these homogeneous constructs and cannot validly measure more complex constructs. Ideally, valid tests aiming to measure complex constructs consist of multiple reliable homogeneous sub-scales or individual tests. Therefore, the ability to which tests can reliably measure the theoretical sub-dimensions is an important indication of its validity.

Scalability

Scalability refers to the degree to which a test-scores are built through a valid calculated of its constituents, the item-scores. In other words, the test-scores must be calculated from the item-scores in such a way that test-participants with a higher capacity on the intended construct also receive a higher test-score. Within classical test theory, test scores are typically created by taking the sum of all item-scores. Within classical test theory, the criterion of scalability is not empirically verifiable.

6.1.3 Steps of test-design.

The development of a psychometric test can be generally divided into four phases: (1) the strategic planning of the test-development, (2) the development of a draft through the systematic formulation of items, (3) the empirical piloting of this draft, and (4) the modification of the draft test based on the analysis of the pilot assessment's results (Bühner, 2011; Jonkitz, Moosbrugger & Brandt, 2012; Rost, 2004).

The first phase of the strategic planning of test-development consists of four steps that are dependent on the theoretical knowledge about the measured construct and the research question. The first step consists of determining the type of indicators. Does the test consist of indicators of subjective (self-evaluations) or objective (achievement)? In a second step, the main characteristics of the population are defined: age, level of educational, language mastery, degree of experience, amount of knowledge and subjective theories. In a third step, a strategy of test-construction should be chosen. The strategy of test construction can follow a rational,

theoretical-deductive logic. Alternatively, it can also follow an inductive exploratory logic, in which a large quantity of items is constructed and the items and the construction of contingencies of items (factors) are selected based on multiple rounds of pilot-assessments. The fourth and last step consists of the definition of the construct and the generating of indicators or test-dimensions. This step aims to define the indicators or theoretical dimensions that form the primary construct. This can be done with a top-down strategy in which indicators are formulated from theory or existing tests. An alternative strategy is a bottom-up, experience-based strategy, in which the questioning of experts leads to the definition of the indicators.

In the second phase, a first draft of the test is produced. There are essentially two kinds of item-formats: natural language, open or *constructed-response items*, in which test-participants are asked to write down or construct the correct answer in naturally occurring language and *selected-response items* in which a test-participant is asked to select the response from a set of alternative responses. Selected-response items consist of *rating-items*, *true-false items*, and *single-*, or *multiple-choice items*. Rating-scales consist of an item-stem with a particular statement followed by a rating scale with which a test-participant can indicate the degree to which the statement applies. With a true-false assignment, test-participants are asked to determine whether a statement is true or false. With single or multiple-choice items, a test-participant is asked to identify which single answer (single-choice) or multiple answers (multiple-choice) represents the correct answer to the question. Constructed-response item formats can include semi-structured *completion items* in which test-participants are asked to fill amend a statement with one or two words that represent the correct answer. Alternatively, *essay items* refer to items where test-participants are asked to write down the correct answer in one or several short sentences.

After the first draft of the test is completed, the test is piloted to identify items that do not meet the criteria for test quality (Jonkitz et al., 2012). Here it is important to make sure that the pilot sample is equivalent to the targeted sample of the main study. Test-piloting is done through an item-analysis and the analysis of test-reliability (Bühner, 2011): Item-analysis consists of the analysis of *item-difficulty*, *item-variance*, *item-total correlation* and *item-reliability* (Jonkitz et al., 2012).

Item-Difficulty (P_i)

The Item-difficulty index indicates the proportion of the total number of test-takers who answered an item correctly (Fishman & Galguera, 2003; Moosbrugger, 2012b). The item-difficulty index provides insight into which items can be considered too easy or too hard. The index is calculated by dividing the average amount of points on an item by the total amount of points that can be scored on that particular item. The index is expressed in *p* of which values lie between 0.00 (none of the test-takers scored a point on that item) and 1.00 (all test-takers received the maximum points on that specific item). A test can be regarded as a suitable instrument for the assessment in a specific sample if the frequency distribution of item-difficulty is normally distributed around the point where most test-participants are expected to score (Fishman & Galguera 2003).

Item-variance ($Var(x_i)$)

Item-variance is an indicator of an item's ability to differentiate between test-takers (Moosbrugger, 2012b, p. 81). When the average variability of an item is high, the item-scores are relatively widely spread-out, and the item can differentiate well between the different test-takers. Item-variance is strongly dependent on item-difficulty (Bühner, 2011). The value of possible item-variance is limited with skewed distributions of item-difficulty. When an item either is very difficult or very easy, this strongly reduces the maximum item-variance and that item's capacity to differentiate between test-takers.

Item-reliability (α_i)

The item-reliability – or item-consistency – index shows for each item to what degree an item contributes to the overall internal consistency of the test (Fishman & Galguera, 2003). Item-consistency is expressed by calculating the statistic *Cronbach's Alpha if item deleted* and is expressed in that same unit (α). This parameter depicts the value of the internal consistency of the test if that specific item would be deleted. Ideally, the values of this index are not below the value of the internal consistency for the test.

Item-Total Correlation (r_{it})

The item-total correlation index shows for each item the correlation between its item-score and the test-score. As such, it is an indication to what degree an item assesses a similar construct as the test as a whole. The item-total-correlation index represents a point-biserial correlation coefficient between the item-score and the test-score and can range between -1.00 to 1.00. Ideally, values on the index are high, showing a strong positive correlation between test-score and item-score. As a rule of thumb, values on the item-discrimination index of .20 can be considered acceptable and values of .40 or higher can be considered good (Moosbrugger & Kelava 2012).

Based on the results of item-analysis and reliability-coefficients, modifications to the test can be made. When an item-analysis shows that certain items show inadequate values, an item can be deleted or reformulated. Typical restructuring occurs because results show that certain items are too easy or too difficult, do not show adequate variance or show low values for item-total correlation. Test-quality can be improved by piloting the draft, evaluating the results of item-analysis and reliability-coefficients, and modifying the test. This process can be repeated multiple times (Fishman & Galguera, 2003).

6.2 CMC Test-Development

The new video test was labeled Classroom Management Competence (CMC). It was aimed to build on the strengths of CME and at the same time to improve CME's theoretical foundation. The following sections discuss the design process of this newly developed video test CMC. Based on the results of research on classroom management, the theoretical framework of CMC is formulated (6.2.1) and the selection-process of video-clips is described (6.2.2). Based on both the theoretical framework and the content of the video-clips, items were formulated (6.2.3). The first version of CMC was piloted (6.2.4), results of item-analysis discussed, and final versions of the items were formulated (6.2.5). For the coding of the answers to the open-response items, a coding-rubric was developed, for which a set of criteria were formulated (6.2.6). The last section describes the coding procedure (6.2.7).

6.2.1 Construct: classroom management competence.

The construct of classroom management competence consists of two dimensions. The first dimension defines one of three classroom management strategies (see chapter 2.3). The second dimension refers to the situation-specific skill that is being assessed (see chapter 3.7).

Classroom Management

The strategies of classroom management are operationalized through the specification of the three foci described in chapter 2.3: (a) a focus on managing group interaction, (b) a focus on managing the effective use of available learning time, and (c) a focus on managing lesson disturbances. Operationalization of these three foci needs to meet three related criteria: Firstly, the operationalization of the theoretical framework should be based on an understanding of classroom management that is relevant to prospective teachers. The strategies should be applicable in practice by novice teachers. Lesson disturbances are of high concern to and cause for stress with starting teachers (Jones, 2006). However, a perspective on classroom management that focuses too much on the ways to react to and deal with lesson disturbances underestimates the effectiveness of a broad spectrum of other teacher behavior: the everyday routines and teacher-student-interactions that are the base of every lesson-situation. Secondly, further operationalization should focus on the specification of strategies that can be trained in a university-based course. Complex dimensions like *withitness* and *overlapping* (Kounin, 1970) that require extensive video-analysis and theoretical reflection are considered less suited than more obvious strategies like for example, the arrangement of the physical classroom (Borich, 2011). Thirdly, the strategies of classroom management that are included in the video test need to be adequately visible in relatively short video clips. Although for example, the teacher-student relationship (Mayr, 2006) or measures of time-management are relevant for classroom management, a video lasting two minutes will probably not provide adequate information to analyze and evaluate the situation. When the above-described criteria to the three foci defined in chapter 2.3 are applied, the construct of Classroom Management Competence consists of the following three strategies (Borich, 2011):

Effective use of the physical Classroom Arrangement (CA)

The physical arrangement of the classroom and the way in which the teacher makes use of the possibilities and limitations of this arrangement strongly influence the teacher's capacity to monitor and influence students' behavior. Moreover, the effectiveness of a specific teaching intervention is dependent on the way students are seated. The strategy of classroom arrangement includes the field of vision of both students and teacher, the positioning of the teacher within the classroom, the position of desks and chairs, learning materials and instructional media. A teacher's effective classroom management is then firstly indicated by his or her ability to:

- arrange the classroom in a way that the intended social form of an exercise can effectively be implemented;
- arrange the classroom in a way that no distractions or obstacles hinder students' learning;
- arrange the classroom so that students can see the teacher;
- arrange media and materials in a way that all students have unobstructed physical and visual access;
- position his or herself in a way that he or she is both visible to the students and he or she can monitor student behavior.

Effectively use of rules and routines (RR)

Rules and routines help to increase the available learning time because they both contribute to the automation of students' behavior. Rules make sure that students know what the behavioral norms are. There are general rules that apply for all lesson situations, for example concerning the use of the bathroom, but there are also rules that apply to specific teaching situations, like for example showing up late for class, participating in a group conversation or early completion of class assignments. General routine can refer to the signaling of the start of a lesson or the transition to a new lesson phase. Routines may also fit to regularly returning specific lesson situations, like for example checking for absentees, checking students' assignments, giving feedback to results. A teacher's effective classroom management is secondly indicated by his or her ability to:

- establish clear rules about appropriate students' behavior in specific situations;

- use routines to signify the beginning of a lesson or transitions between lesson phases;
- use routines to instruct students about assignments.

Effective use of low-profile classroom management techniques (LP)

Low-profile management refers to the set of management strategies that aim at keeping disturbance of the lesson flow minimized. Low-Profile techniques firstly include alertness for changes in student behavior, motivation, attentiveness or excitability. Secondly, it includes the deflection of lesson disturbances through the use of nonverbal cues to react to classroom situations, even before they have developed into a disturbance. Thirdly, when disruptive behavior occurs and a teacher-intervention is required, low-profile management includes communicating clear expectations of student behavior, stating consequences and postponing discussions to after the lesson. A teacher's effective classroom management is thirdly indicated by his or her ability to:

- quickly identify and react to situations that might result in lesson disturbances;
- choose an intervention that does not interfere with the lesson flow;
- choose an intervention that does not disturb other students from their work;
- minimize the showing of anger or frustration;
- choose an intervention that is short, so that loss of instructional learning time is minimized;
- choose an intervention that does not expose or embarrass individual students.

Situation-specific skills

The dimension of competence is included in the operationalization of CMC through a specification of the situation-specific skills discussed in chapter three. The competence aspect of CMC is strongly oriented on the three cognitive demands defined for CME and the research on situation-specific skills discussed in chapter three: *Accuracy of Perception*, *Holistic Perception,* and *Justification of Action*. (König, 2015b; König & Lebens, 2012).

Accuracy of Perception (AP)

Accuracy of perception refers to the degree to which a test-participant can identify relevant

behavior. It is assumed that a test-participant with a high degree of classroom management competence applies a filter that allows for a quick distinction between relevant and irrelevant teaching aspects. Consequently, competent test-participants can commemorate certain aspects of a video clip with a high degree of precision because their filter made them focus on these aspects.

Holistic Perception (HP)

Holistic perception refers to the degree to which a test-participant is firstly able to place a specific event into its timely and spatial context within the depicted classroom episode. Secondly, holistic perception refers to the degree to which a test-participant can view this event as an example of a typical aspect of teaching. It is assumed that a competent test-participant is firstly able to relate a specific event to other events occurring before or after the event in question and is able to discern cause and effect relationships between them. Secondly, a competent test-participant can formulate likely alternatives to teacher action. Thirdly, a competent test-participant can name events seen in the video clip that form an example of a more general category or statement.

Justification of Action (JA)

Justification of action refers to the degree to which a test-participant is able to infer the likely reasons for a specific teacher's behavior or intervention. This means that it is assumed that a competent test-participant can firstly derive the probable intentions of a teacher within a specific situation. Secondly, a competent test-participant can mention the advantages of a specific teacher's intervention. Thirdly, a competence test-participant is able to name probable causes of a specific teacher action.

These three cognitive challenges refer to three levels of competence. AP represents the most concrete level as test-participants are asked to precisely reproduce an event from their memories. HP represents the first level of inference, where test-participants are asked to relate events to other events and classify a specific event as an example of a more abstract phenomenon. JA represents the second level of inference, where test-participants are asked to interpret an event regarding the intentions or reasons of a teacher. AP and HP are considered to represent the situation-specific skill *perception*. JA is considered to represent *Interpretation*. The

situation-specific skill *decision-making* is not operationalized in the construct of CMC.

6.2.2 Selection of video clips.

For CMC, four new video clips were identified. The following criteria were formulated for the process of selection of video clips. Firstly, video and audio needed to be of high quality. Secondly, mirroring the teacher training program of the sample, video clips should depict teaching from secondary education. Thirdly, the scenes should portray teachers that are heavily challenged in keeping order and that show authentic and typical classroom management behavior, but not necessarily either best or bad practice. Fourthly, to reduce learning effects within the test, the four videos should depict four different teachers. Fifthly, video clips should depict a whole (sub-) segment of classroom interaction and should not break into an existing conversation or cut-off a segment in the middle. Sixthly, video clips were aimed to last around two minutes. Seventhly, video clips should depict all of the classroom management strategies: the physical arrangement of the classroom, the establishment of rules and routines and the use of low-profile management strategies.

The video clips were selected from the video corpus of Professor Dr. Udo Rauin, recorded as part of the research project called *Strategien des Unterrichts in heterogenen Klassen und ihre Wirkung auf Schüleraktivität* (Appel, 2016). The selection process consisted of four steps. In a first step, 17 lessons of five different teachers were examined and divided into lesson segments (Dinkelaker & Herrle, 2009). The second step consisted of evaluating the relevance of each of the segments to the subject of classroom management. In a third step, the segments with a high-level relevance to classroom management were examined in more detail by applying the above-formulated criteria. As a result, six possible video clips of four different teachers were identified. In a fourth step, the final four video clips were selected and their exact timely boundaries were defined. After completing the above-described process, the following four video clips were selected to be used in the test.

1. “Boys and Girls, listen to me” (02:05)

The first video shows a female teacher initiating a transition to a new phase. Students are working

on a group assignment. The teacher aims to summarize the results of the assignment and to evaluate them on the blackboard. She tries to attract student's attention by positioning herself centrally, raising her voice, and using different verbal and non-verbal measures to attract students' attention. As this does not produce the intended effect, she turns to pausing and waiting to attract students' attention. When all students direct their attention to her, she gives the instruction that one student of each group should hang the working sheets on the blackboard. When all students return to their seats, students are chattering, and the teacher needs to attract their attention a second time to give feedback on the working sheets. Running out of time and patience, she then calls out one student who is sitting with her back to the teacher and issues her to turn around. The student turns around, which causes all students to direct attention to her. The teacher then starts giving feedback on the exercise.

2. "First of all: Be Quiet and Listen" (02:06)

The second video shows a female teacher before the start of her lesson. Almost all students sit at their desks, which are arranged in groups. The teacher moves through the classroom and removes a water bottle that is still standing on one of the desks and proceeds to ask a student to move away from his neighbor. She shortly waits to attract attention and then starts the lesson by asking the class about the date that was written on the blackboard beforehand by one of the students. Some students start calling out the answer. She ignores these call-outs, tells one student that it is not her turn and asks another student (Peter) for the answer. He does not know the answer, so she asks another student (John) who says the date correctly. Directly after this, the teacher turns back to Peter and asks him a second time for the date. Peter cannot give the correct answer. Agitatedly, the teacher points out to Peter that John just mentioned the correct answer and Peter did not pay attention. She then asks a third student for the date.

3. "Good Morning to You" (01:58)

The third video shows a female teacher standing close to the doorway before the start of the lesson as the students are walking towards their seats. Seats and desks are arranged in forward-facing pairs. She has several short conversations in which she reminds students to take their seats and spit out chewing gum. She then moves towards her desk to look at the seating plan and

comments to one student that the seating arrangement has been changed since the last lesson. She then tells all students to stand up and start the lesson like they always do. After reminding several students to stand up, all students are standing. She initiates a greeting ritual by wishing all students good morning. Students reply by singing a welcoming song. After completion of the song, all students sit in their chairs and face the teacher.

4. “3, 2, 1 and stop”, (02:03)

The fourth video shows a male teacher as he instructs students to stop working on their assignments and pay attention to his instructions by calling out “3, 2, 1 and stop!”. The teacher continues by giving instructions on the next assignment, during which students should listen to an audiotape while simultaneously reading the text in their textbook and whispering the words. The teacher reminds the students several times to whisper and turns on the audiotape. While students are working, he scans the classroom. He notices one girl, walks towards her while agitatedly saying, “what’s wrong with you!?” The student then puts non-lesson related materials that were distracting her. The teacher gives her a disapproving look while asking her if she underlined bits of the text in her textbook. She says she did not, after which the teacher returns to his desk.

6.2.3 Operationalization and item-development.

The construct of classroom management competence was operationalized following the four phases described by Markus Bühner (Bühner, 2011) and discussed in chapter 6.2. The first phase consists of the strategic planning of the video test. CMC is considered an objective achievement test, measuring situation-specific skills in classroom management. The population consists of the upper secondary education teacher training track, indicating a minimum age of 19 and good to very good language capabilities. Because the population is expected to fall in the first four terms of their teacher training program, knowledge, subjective theories and degree of practical experience are expected to be novice-like. The construction of the test is heavily dependent on the content of the video clips. As such, the number of possible items is limited and consequently, an inductive exploratory strategy would not be appropriate. Instead, a theoretical-deductive strategy was used, in which every single item firstly represents one of the three classroom

management strategies - *Classroom Arrangement*, *Rules and Routines*, and *Low-Profile* - and secondly, one of the three competence-related cognitive demands - *Accuracy of Perception*, *Holistic Perception*, and *Justification of Action*. It was aimed to distribute the items evenly across all three classroom management strategies and cognitive demands.

The second phase consists of the development of a first version of the test. Following the steps for the strategic planning described above, 30 items were developed for the first version of the test. Of these 30 items, 19 items were formatted as open-response and 11 were selected-response. The open-response items are all formatted as essay items, in which test-participants are asked to write down their answers in one or several short sentences. Of the 11 selected-response items, 10 were single-choice items and one item consisted of five true-false items. The distribution of the items across the three classroom management strategies and the three cognitive demands is presented in Table 1.

Table 1
Distribution of Items of Preliminary Version of CMC

Theoretical dimensions	Classroom Arrangement (CA)	Rules and Routines (RR)	Low Profile (LP)	Total
Accuracy of Perception (AP)	4	4	5	13
Holistic Perception (HP)	2	2	3	7
Justification of Action (JA)	3	3	4	10
Total	9	9	12	30

Eight example items, used in the first preliminary version of CMC, are listed in Table 2 in their English translations[4]. Appendix A lists all the items as they were used in the preliminary version of CMC in their original German wording.

.

[4] German translations by author.

Table 2
Examples of Items, Theoretical Dimensions, and Item-Format of CMC (preliminary version)

Video Clip	Item	CM Dimension	Situation Specific Skill	Item Format
1.	Please name the techniques used by the teacher that are aimed at gaining her students' attention.	LP	AP	Essay-Item
1.	Which employed technique proved to be especially effective?	LP	HP	Essay-Item
2.	Why does the teacher interrupt her introduction of the lesson and move towards the other side of the classroom?	RR	JA	Single-Choice
2.	The teacher asks one student twice about today's date. What does the teacher aim to communicate by asking him a second time?	LP	JA	Essay-Item
3.	What could be the reason that the teacher asks all students to stand up before she starts with the classroom ritualized greeting?	RR	JA	Essay-Item
3.	Please name the effects of the ritualized greeting on the student behavior	RR	HP	Essay-Item
4.	Which of the following statements fits best to the classroom arrangement? - In the front row, only boys are seated. - In the front row, only girls are seated. - In the front row, both girls and boys are mixed.	CA	AP	Single-Choice
4.	What could be the reason that the teacher moves towards the student, before reprimanding her?	CA	JA	Essay-Item

Note. AP = Accuracy of Perception; HP = Holistic perception; JA = Justification of Action; PA = Physical Arrangement; RR = Rules and Routines; LP = Low Profile;

6.2.4 Pilot-assessment.

A pilot-assessment of the CMC video test was held for four reasons. Firstly, to evaluate the practical feasibility of an alternative technical setup. Secondly, to evaluate the relevance and authenticity of the videos and items from the perspective of the test-participants. Thirdly, to evaluate the coding procedure that was performed in specially formatted spreadsheets. Fourthly, to evaluate criteria for test quality so that a specific improvement to item-formulation could be made.

Pilot sample

CMC was piloted in April 2016 in the second session of a course on the subject of classroom management that consists of a total of 15 sessions. A total of n = 29 prospective teachers participated in the test. Test-takers were 72.4% female. Their mean age was M = 28.6, SD = 8.0. The youngest participant was 19 and the oldest 42. Test-takers were enrolled in four programs

of teacher training: 1 in primary education, 13 in lower secondary education, 12 in higher secondary education, and 3 in special needs education. Test takers were in different stages of their studies. The mean number of enrolled semesters was *M* = 4.5; *SD* = 2.5. The minimum of enrolled semesters was 2, the maximum 11. Test participants were in different stages of the completion of their two university-based internships. 12 (37.9%) of the test-takers had completed the first short internship. Of this group, seven prospective teachers (20.7%) had also finished the second internship.

The sample of the pilot study consists of only prospective teachers that are in different stages of completion of their training programs and internships. However, there several specific points where the pilot sample does not represent the target population (see chapter 6.6). Firstly, the pilot sample consisted of prospective teachers of all teacher training programs, whereas the population exclusively consists of prospective teachers within the higher secondary program. Secondly, the pilot sample is 72.4% female, whereas the target population is roughly 60% female (Goethe University Frankfurt). Thirdly, the mean number of enrolled semesters was 4.5 in the pilot sample, whereas the target population consists of prospective teachers in either their second or fourth term (see chapter 6.6). It is concluded that the sample of the pilot study can be considered roughly representative, but does not exactly represent the population of the main study.

Test-administration

Instead of a paper-pencil test, the technical setup of the evaluation project of the practical semester was used (Practical Semester Hesse, 2018). Questionnaires were presented and filled-out on Android-based tablet computers with a browser-based app that was developed using the open-source software Lime-Survey (LimeSurvey, 2016). Participants typed their answers on a wireless keyboard connected to their tablet computers through Bluetooth. The tablet computers were connected to a local Wi-Fi network that was hosted by a laptop. The participants' responses were transferred wirelessly and in real-time from the tablets to the LimeSurvey-server that was administered on the laptop. After the completion of the test, test-entries were downloaded from the LimeSurvey server and exported in a *.dat* file-format. These files were imported in SPSS 23

for further analysis. 12 Rating-items from the Observer evaluation study were added (Seidel et al., 2010) to evaluate test participants' ratings of the videos and items.

Before the test participants entered the course room, the technical setup was prepared and tested. After the participants had taken their seats, test instructions were read and participants were informed that participation was voluntary and anonymous. The test started with the introductory items. After this first phase, participants collectively watched the first video clip once, without the possibility to pause or replay, on a centrally placed screen in the course room. With the ending of this first video, a timer was set for eight minutes. In CME, test-participants have 7 minutes to answer 24 questions. Because the pilot sample consists of 30 items, it was decided to set the timer for 8 minutes. A prompt was given one minute before the eight minutes were expired. As was expected, almost all test-participators finished answering the items within that time-frame. Directly after the eight minutes were expired, the second video was played. This process was repeated for the third and fourth videos. After the time of the fourth video clip was expired, test-participants were prompted to skip to the last section of the questionnaire, in which they were prompted to answer the rating-items related to the evaluation of the videos and the test. During the test-administration, minor connection issues occurred on a handful of tablet computers that needed to be resolved before the start of the test. Other than these minor issues that did not further interfere with the assessment, the utilized alternative setup worked flawlessly: The laptop which acted as a local server successfully gathered data from all test participants.

User-ratings of videos and items

Test-participants rated each item separately for its comprehensibility on a 4-point Likert scale (1 = hardly comprehensible, 4 = perfectly comprehensible). The mean rating for all 30 items was M = 2.2 and the average standard deviation for all items was SD = 0.38. These values indicate that on average, the test-participants rated the items' comprehensibility as mediocre. Moreover, mean ratings per item did not vary much (M_{min} = 2.03; M_{max} = 2.34), indicating that there were no clear items that were rated either as very comprehensible or very incomprehensible. Additionally, test-participants rated the items and video clips as a whole on a 4-point Likert scale

(1 = doesn't apply, 4 = applies). Table 3 reports the results of the participants' general ratings of CMC's video clips and items using the items of (Seidel et al., 2010, pp. 303–304).

Table 3

Means and Standard Deviations of Pilot-Rating of Preliminary Version of CMC[5]

Item	*M*	*SD*
"I thought the video clips were..."		
...informative	3.25	0.75
...too short	1.54	0.79
...untypical	1.46	0.69
...authentic	3.34	0.67
...interesting	3.41	0.73
...diverse	3.00	0.90
"I thought the questions to the video clips were..."		
...hard	1.75	0.84
...interesting	3.17	0.81
...appropriate	3.31	0.71
...inappropriate	1.46	0.84
...too lengthy	1.43	0.63
...diverse	3.00	0.89

These results show that the prospective teachers in the pilot sample generally rated the video clips of CMC to be, informative (*M* = 3.25), authentic (*M* = 3.34), interesting (*M* = 3.41), diverse (*M* = 3.00), and to be of appropriate length (*M* = 1.54). In contrast to the individual item-ratings, test-participants rated the items to be not too hard (*M* = 1.75) nor too lengthy (*M* = 1.43), while at the same time being interesting (*M* = 3.17), appropriate (*M* = 3.31) and diverse (*M* = 3.00). These last values indicate that, on average, test-participants in the pilot sample rated the video clips and items as being appropriate for use in the video test. The results of both user-ratings seem to contradict. The comprehensibility is rated as mediocre, whereas the items are not rated as too hard, lengthy, or inappropriate. It was concluded that items needed to be evaluated and if necessary, rephrased for comprehensibility.

Coding-procedure

Before the data could be coded, the answers to the essay-items were copied in spreadsheets that

[5] English translation by the author.

were specifically formatted for coding in the software Excel 2013. For the coding procedure, several criteria were formulated for each open-response item. The structure of the coding rubric is identical to that of CME: for each item, one or more criteria were formulated. Each of the criteria represented a correct answer to the question. These criteria were formulated on the basis of theories on classroom management by the author of the current study. In a first round, each criterion is coded dichotomously (1 = criterion met, 0 = criterion not met). Because of restrictions on recourses, only one trained rater coded the answers of the open-response essay-items. Dichotomous codes for each of the criteria were directly entered into the designated cells of the coding sheet. To create scores for each item, these dichotomous codes were coded in a second round, where a point was given when at least one-third of the total amount of criteria was met. Based on the analysis of item-difficulty, the conditions for gaining an item-score in this second round were reviewed for each item. When an item was considered to be too easy, more criteria were needed to gain a score on the item. When an item was considered to be too difficult, fewer criteria were needed to gain a score on the item. Answers to the single-choice items could be derived from the exact observation of the video clip. The item-scores were imported into SPSS 23 for further data analysis.

Internal consistency and item-analysis

To empirically evaluate the quality of the draft of CMC, an analysis of internal consistency and an item-analysis were performed using SPSS 23. Table 4 shows the results of the item-analysis.

Table 4

Item-Analysis of Assessment of preliminary version of CMC

ID	Situation-Specific skill	CM-Dimension	Format	*P*	Var(x_i)	Alpha if item deleted	r_{it}	Item-Changes
1.1	AP	LP	OR	.45	.26	.51	.49	no
1.2	HP	LP	OR	.24	.19	.53	.43	no
1.3	AP	CA	SR	.62	.24	.61	-,21	small
1.4	AP	LP	SR	.59	.25	.54	.34	no
1.5	JA	LP	OR	.07	.07	.58	-.04	big
1.6	JA	CA	OR	.97	.03	.56	.34	big
1.7	HP	LP	SR	.34	.23	.53	.39	no
1.8	HP	LP	OR	.72	.21	.62	-.34	big
2.1	JA	CA	SR	.97	.03	.56	.22	discarded
2.2	AP	CA	OR	1.00	.00	.57	.00	big
2.3	JA	RR	OR	.07	.07	.56	.22	small
2.4	AP	CA	SR	.17	.15	.56	.18	big
2.5	AP	LP	OR	.90	.10	.56	.14	small
2.6	JA	CA	OR	.14	.12	.59	-.14	discarded
2.7	JA	CA	OR	.76	.19	.55	.23	small
2.8	AP	RR	SR	.79	.17	.59	-.12	big
3.1	AP	LP	SR	.97	.03	.59	-.26	big
3.2	JA	RR	OR	.59	.25	.56	.17	no
3.3	HP	CA	OR	.93	.07	.55	.32	big
3.4	AP	RR	SR	.66	.23	.58	.07	small
3.5	JA	RR	OR	.14	.12	.58	.01	small
3.6	AP	RR	SR	.72	.21	.55	.24	big
3.7	HP	CA	OR	.28	.21	.53	.38	small
3.8	HP	RR	OR	.24	.19	.55	.29	no
4.1	AP	RR	OR	.31	.22	.51	.53	no
4.2	AP	LP	OR	.55	.26	.52	.42	no
4.3	HP	RR	OR	.03	.03	.58	-,.21	discarded
4.4	AP	CA	SR	.48	.26	.53	.40	no
4.5	JA	CA	SR	.72	.21	.58	.02	big
4.6	JA	CA	SR	.14	.12	.59	-.14	discarded

Note. AP = Accuracy of Perception; HP = Holistic perception; JA = Justification of Action; PA = Physical Arrangement; RR = Rules and Routines; LP = Low Profile; P = Item-Difficulty; Var(x_i) = Item-variance; r_{it} = Item-total Correlation.

Analysis of internal consistency for the pilot study showed a result of $\alpha = .57$, which can be judged as poor (Fisseni, 2004). However, item-analysis also showed that some of the items were either too hard or very easy. Moreover, the negative values for corrected item-total correlation on some items indicated that test-participants interpreted the item falsely, for example, because of an unclear formulation, or because the item in some other way did not represent the intended construct of classroom management competence. After discarding these items from the analysis, 18 items remained for which the internal validity showed a result of $\alpha = .77$, which can be considered an acceptable value (Fisseni, 2004). On the basis of these results, it was decided to leave the core of 18 items and reformulate the items that showed poor psychometric values.

Based on the results of item-analysis, each of the items was evaluated and reasons for possible poor values were inferred. There could be different reasons for poor values on item analysis: The item-wording could be unclear, too complex, or too long or the answer to the question might have been too easy or too hard. Depending on the results of the evaluation, some of the items were reformulated. Table 4 shows the degree to which the item-formulations changed. 'No' stands for items that stayed exactly like they were formulated in the pilot study. 'Small' refers to relatively minor item-changes, for example in the clarity of their wording, or through the addition of an answering-category. 'Big' refers to relatively major item-changes that leave the topic of the item intact, but for example, change the item-format (single-choice to open-response) or strongly modify the item-formulation. 'Discarded' refers to the items that were eliminated. In total, four new items were formulated, based on the guidelines discussed in chapter 6.2.3 and the insights from item-analysis discussed in this section. The final version of CMC consists of a total of 30 items. This version of CMC was used for the main assessment. Appendix B lists all the items of the final version of CMC.

Evaluation of results of pilot-assessment

As stated in the introduction, the pilot assessment was held for four reasons: to evaluate the practical feasibility of an alternative technical setup, to evaluate relevance and authenticity of the videos and items from the perspective of the test-participants, to evaluate the coding procedure that was performed in specially formatted spreadsheets, and to evaluate criteria for

test quality so that specific improvement to item-formulation could be made. Each of these objectives is discussed below. The pilot assessment showed that the alternative setup, in which participants fill out the questionnaire on tablet computers, poses a viable alternative to using paper-pencil tests. The technical setup worked without noteworthy problems and data could be successfully stored and exported from the LimeSurvey server.

Test-participants rated the video clips as informative, authentic and interesting. In general, the items were rated as being appropriate, diverse and neither too lengthy nor too hard. However, separate ratings of each item indicated mediocre comprehensibility. Results of both user ratings, the ratings for the test as a whole (see Table 3) and the ratings of the individual items, contradict each other. In sum, it was concluded that both the videos and the items could be considered adequate for the use within a video test, but that revision of each of the items should be done with extra care on clarity and comprehensibleness.

The coding of participants' answers to open-response items within spreadsheets could be conducted objectively and quickly. Because only one rater did coding, inter-rater reliability could not be calculated. As a result, the quality of the coding procedure should be interpreted with reservations. Reported values of Inter-Rater-Reliability of CME (König, 2015b) can be considered high (mean $\kappa > .8$.) Because the item-structure and coding procedure of CMC are similar to CME, and coding was performed by a rater who had already gained experience with the coding of CME, it is assumed that coding quality was at least satisfactory.

Calculation of Internal consistency and item-analysis showed that some items were either too hard, too easy or in some other way needed revision. Evaluation of the results of item-analysis led to the following changes: nine Items stayed the same, seven items underwent small changes in their wording. Ten Items underwent bigger changes in their wording or item-format. Four items were discarded from the test entirely and replaced by new items.

6.2.5 Overview of all items.

The final version of CMC consists of 30 items, of which 22 are open-response items, six are single-choice items and two are true-false items. Table 5 gives an overview of the format and

dimensions of all items. Appendix B lists all items that were formulated after the revisions resulting from the pilot study of the preliminary version.

Table 5
Distribution of Items of Final Version of CMC

Theoretical Dimensions	Classroom Arrangement (CA)	Rules and Routines (RR)	Low Profile (LP)	Total
Accuracy of Perception (AP)	3	4	5	12
Holistic Perception (HP)	2	4	3	9
Justification of Action (JA)	3	3	4	9
Total	8	10	12	30

6.2.6 Development of coding rubric.

For the final version of CMC, a coding rubric was developed using a method that is similar to the method of CME. Coding consisted of two phases. In the first phase, all answers to the open-response questions were coded with the help of criteria. Each criterion represents a possible correct answer to the question. Whenever a criterion applies to the answer as formulated by the test-participant, a 1 is coded for this criterion. When a criterion does not apply, a 0 is coded. When a question was not answered, all criteria for that question were coded with an 8 for 'conscious omission'. When a question was not answered and no further questions for this video were answered, all criteria for this question were coded with a 9 for 'not reached'. All codes are re-coded in the second phase. In this second phase, the relative weight of all the criteria was taken in to calculate an item-score. Because an omission was interpreted as being a conscious decision to not answer the question these codes were recoded with the item-score 0. When a question was not reached within the available time, the corresponding codes of the first round were recoded as missing.

This coding rubric for the first round of coding was developed on the basis of a deductive (theory on classroom management) and inductive (the answers provided by test-participants) process. The inductive process consists of three rounds of expert-ratings. The deductive process involves the formulation of examples of correct and incorrect answers to each item.

Formulation of criteria for the open-response items

The open-response items were coded by applying a set of criteria for each answer. Each criterion represents a possible correct answer to the question posed in the item. The set of criteria was formulated throughout two rounds of expert-ratings. In the first round, four teacher-trainer experts and the author of the current study watched the videos and independently formulated all possible correct answers to the open-response questions. All four experts had experience with teaching the topic of classroom management for prospective teachers. Their teaching experience ranged from 2 to 6 terms. The author of the current study collected and summarized all answers and reformulated them in model answers. For all 22 open-response questions, 106 model answers were formulated. Table 6 lists the criteria for the example essay-items listed in Table 2. Two items listed in Table 2 are not included in Table 6, because they were single-choice items for which no coding criteria were constructed. A complete list of all criteria and their expert-ratings are listed in Appendix B.

In the second round of expert-ratings, a different group of teacher training experts consisting of 18 in-service teachers participated in a questionnaire. These in-service teachers have multiple years of teaching experience in schools and additionally taught in one or more university-based courses related to the mentoring of the internships. As such, these expert teachers were assumed to have both theoretical (from their mentoring of internships) as well as practical (from their own teaching experience) expertise. All experts were asked to watch the video clips, inspect the items and rate to what degree the 92 model answers represent a correct answer to the question on a scale of 1 to 5. Results of the expert-ratings of all 92 model answers show a mean score of M = 4.02 (SD = 0.73). Four model answers with a mean rating lower than 2.6 were discarded from the analysis. The formulation of one model answer with a mean rating of 2.6 was changed because its formulation was, in retrospect, considered unclear. The remaining 87 criteria had a mean rating of M = 4.12; SD = 0.64.

Table 6
Examples of Item-Wording and Criteria of CMC.

Video Clip	Item	Criteria
1.	Please name the techniques used by the teacher that are aimed at gaining her students' attention.	- Change of teacher's position within the classroom - Express explicit expectations of student behavior (listen to me!) - 'Shh'-noises - Gesticulation - Being quiet, breaking off sentences - Stating the student's name
1.	Which employed technique proved to be especially effective?	- Stating the student's name - Being quiet, breaking off sentences
2.	The teacher asks one student twice about today's date. What does the teacher aim to communicate, by asking him a second time?	- The student should pay better attention - To show interest in his learning result
3.	What could be the reason that the teacher asks all students to stand up before she starts with the classroom ritualized greeting?	- As an expression of respect and politeness - To get a quick overview, who is on-task - To underline equality among students - To communicate that the lesson is starting
3.	Please name the effects of the ritualized greeting on the student behavior	- All students sit at their desk - All students are quiet - The teacher has the attention of the students
4.	What could be the reason that the teacher moves towards the student, before reprimanding her?	- No distraction of other students - Minimizing exposure of the student - To be better able to evaluate the situation - Physical proximity communicates dominance

Building of item-score

In the next step, dichotomous codes for each criterion were transformed into a trichotomous (0-1-2) item-score. In CME, item-scores are constructed by adding up all the scores of the criteria belonging to one item. If this value surpasses a predetermined threshold, the score for that item becomes 1. If this threshold is not surpassed, the item-score is 0. This implies that within CME, all criteria are assumed to be equally important to the construct of classroom management. For CMC, an alternative systematic was chosen to construct an item-score. It was assumed that although all criteria represent valid answers in themselves, some criteria represent a better answer to the question than others. In other words, receiving a score on one criterion indicates

a higher degree of competence than receiving a score on another criterion. Therefore, counting the quantity of all criteria and defining the threshold does not necessarily represent a good indicator of competence. The third round of expert-ratings was held with two university teachers and the author of the current study with the aim of distinguishing between the criteria with high relevance for the construct of classroom management competence and the criteria with a lower relevance. The two researchers have a minimum of six terms of experience in the teaching of classroom management and video-based analysis and hold a Ph.D. in pedagogy. In a group discussion, the experts were asked to rate the relevance of each criterion for the construct of classroom management competence. Criteria that were rated to be of low relevance were coded with a 1. Criteria that were rated to be of high relevance were coded with a 2. Appendix B presents an overview of these expert-ratings.

Item-scores were firstly based on the results of this third round of expert-ratings. A participant would only receive an item-score of '2' if he or she had written down an answer that was coded with a criterion that was rated as highly relevant in the third round of experts. When a participant's answer only met the criteria that were rated as less relevant in the third round of expert-ratings, the item-score of his or her answer would not exceed 1, independently on the amount of less relevant criteria. The exact thresholds for the item-scoring procedure were based, secondly, on the frequencies of each criterion as coded within the main study. To determine these frequencies, an item-analysis on the criterion-level was performed right after coding of the results of the main study. Additionally, as a result of this item-analysis, a total of 15 criteria were excluded from further analysis, because their frequency and resulting item-variance were very low (Var (x_i) < .05). Thus, the calculation of Item-scores was the result of a deductive and inductive process. Inductively, item-scores were based on the theory of classroom management (a total of three rounds of expert-ratings). Deductively, item-scores were based on the frequencies of criteria as the results of the codes of the main assessment.

Coding of selected-response items

The answers to the selected-response items could be derived by closely observing the video clips. Consequently, no experts needed to be questioned to formulate the correct answers. Instead,

the author defined the correct answer to the selected-response items.

6.2.7 Coding procedure.

For the coding of the open-response items, a coding manual was developed following the guidelines of quantitative content analysis (Döring & Bortz, 2016b) and the structure of the coding manual of CME. In a first deductive step, each criterion was given a title and code, after which the model answer for each criterion and several options for equivalent answers were written out. In a second inductive step, four independent raters applied the coding manual to the questionnaires of 10 test-participants that were randomly selected from the main assessment. Raters wrote down examples of good answers that were considered equivalent to the model answer. They also wrote down examples of wrong answers, for example, because they were considered too vague. When necessary, they formulated proposals for amendments to the coding guidelines. These proposals were discussed within the group and the group made decisions for the adaptations or amendments of the coding manual. This process was repeated a second time with all four raters and a third and fourth time with only two independent raters.

The use of formatted spreadsheets for coding was successfully evaluated in the pilot study (see chapter 6.2). Therefore, these coding-sheets were adapted to be used in the final version of CMC. A spreadsheet document was created in which each sheet represented all relevant information of one test-participant: a list of all the items, followed by a blank column in which the answer to that question would be imported. This column was followed by several rows of cells, in which each cell represented one criterion. Data were extracted from the lime-survey export file and pasted into the empty cells.

CMC was coded by the two raters who also piloted and adapted the coding manual of CMC. Training consisted of two rounds of parallel and independent coding of the answers of 10 test-participants, comparing and discussing the differences between the two coders. Because the two raters were also involved in the multiple rounds of inductive adaptation of the coding manual (see chapter 6.2.8), relatively high degrees of percent agreement could be reached during training.

All data were independently and parallel coded by both raters. If the criterion applied to the answer that was provided by the test-participant a 1 was entered into the respective cell of the coding-sheet. If the criterion did not apply a 0 was entered. When a test-participant did not fill out an answer to the item, an 8 was entered for all criteria on that item. The multiple-choice items were coded automatically according to the guidelines of the coding manual by applying a pre-formulated formula. After the two raters completely coded all data, it was determined on which criteria the codes of the two parallel raters differed. Agreement on these different codes was then reached through discussion between the two raters.

After reaching an agreement on all codes, a second round of coding was conducted to build item-scores on the basis of the codes on the criteria. This second round of coding was performed on the basis of the rating of the third round of expert-questioning described in chapter 6.2.6. When a test-participant mentioned at least one criterion that was rated as being highly relevant for the construct of classroom management competence, that item was scored with a 2. When a test-participant mentioned only criteria that were rated as being less relevant for the construct of classroom management competence, that item was scored with a 1. Adaptations to the systematic of item-scoring were made on the basis of item-analysis (see chapter 7.1).

6.3 Treatment

Based on the general principles of the situated and cognitive strategy (see chapter 4.3), two treatments were designed. This chapter describes the setup of the courses in the current study. In the first section, the institutional context of Goethe University Frankfurt is described (6.3.1). Here, the way in which the course was institutionally integrated and the content of the practical semester is described. In the second section, the general design guidelines of both courses are described (6.3.2). In the third section, the systematical differences between the two courses are described (6.3.3).

6.3.1 Teacher training at Goethe University Frankfurt.

In total, over 6500 prospective teachers are registered within one of the teacher training programs at Goethe University Frankfurt. Goethe University offers teacher training programs for

primary education, the upper and lower level of secondary education, special needs education, and vocational training. The intervention was held as part of the program for the upper level of secondary education called 'Gymnasium'. During the time of the evaluation, over 2500 prospective teachers were registered for the Gymnasium teacher training program (Goethe University Frankfurt). The curriculum of the Gymnasium teacher training program consists of 240 ECTS, spanning a period of eight semesters. The program consists of four parts: two teaching subjects, a practical semester, and 'Bildungswissenschaften', in which subject unrelated general pedagogical knowledge is conveyed.

Practical semester

In March 2013, the Hesse ministry of culture decided to initiate a pilot project that encompassed the substitution of the two separate internships each lasting five weeks with one single internship lasting one semester: the practical semester. The pilot project was divided across three Hesse universities, each of these piloting the practical semester for one or two of their teacher training programs. Goethe University was assigned to realize the practical semester within the gymnasium teacher training program for prospective teachers in the higher level of secondary education. Prospective teachers participate in the practical semester in either the third or fourth semester of their teacher training program. This semester represents a full-time work-load of 30 ECTS. During the practical semester, prospective teachers visit their school for four days a week and their university for one day per week during a period of four months (Practical Semester Hesse, 2017).

During the practical semester, prospective teachers are required to participate for a minimum of five hours per day in a multitude of activities, like the observations of other teachers' lessons and teaching themselves. Prospective teachers are required to teach a minimum of 16 lessons. Moreover, prospective teachers are required to participate in other teaching-related activities like for example, parent-teacher interviews, school excursions, conferences, and school projects. Each prospective teacher has a mentor-teacher in school that functions as a contact person, helps him or her in the preparation of and reflection on their lessons, and introduces him or her to the rest of the school and its extracurricular activities.

One day per week, prospective teachers visit university and attend two courses: one course related to the teaching methods of one of their teaching subjects and one course that focuses on subject-independent aspects of general pedagogical knowledge. Both courses are led by an experienced teacher that is employed by the university and aims at the theoretical reflection of the practical experiences acquired in school. The course related to the teaching subject mainly focuses on subject-specific issues like teaching methodology and the preparation and use of learning materials. In the course on general pedagogy, general aspects of teaching quality, the role of the teacher, personal reflection strategies, and the topic of inclusion are addressed. The theme of classroom management and dealing with difficult students is addressed in at least one session of the subject-independent course. During their internships, prospective teachers are visited at least two times by each of their university teachers and are given feedback on their teaching. As part of these university courses, prospective teachers firstly write a portfolio in which practical experiences are written down and reflected. At the end of the practical semester, the university teacher and the prospective teacher fill out an assessment tool called FIT-L (Schaarschmidt, Kieschke, & Fischer, 2017) and hold a reflection interview. Thirdly, prospective teachers write an internship report in either the course on teaching methods or the course on general didactics. In this report, prospective teachers focus on one aspect of their internship and reflect this aspect with scientific literature. Only the internship report is graded.

The practical semester pilot project also involves an extensive research program. The aim of this project is the scientific evaluation of the effects of the old internship structure and the pilot project across all three involved universities (Practical Semester Hesse, 2018). As part of this research program, all prospective teachers are questioned three times: in the first week, in the fifth week and the last week of their practical semester. The questionnaire involves a multitude of scales including, for example, scales on motivation, self-assessment, and personality.

Bildungswissenschaften

Prospective teachers of Goethe University Frankfurt can partake in specialized courses on the subject of classroom management. Classroom management is part of the curriculum called *Bildungswissenschaften* that consists of general, subject-unrelated didactics.

Bildungswissenschaften consists of four modules named *Unterrichten* (teaching), *Erziehen* (socialization), *Beurteilen* (diagnostics), and *Innovieren* (teacher profession and school development) that are located within the domain of general pedagogical knowledge see (Baumert & Kunter, 2013). Courses on classroom management are located within the module Unterrichten (teaching) (Terhart, Schulze-Stocker, Holzberger, & Kunina-Habenicht, 2013). In programs of teacher education at Goethe University Frankfurt, the topic of classroom management is not mandatory.

The courses that were evaluated as part of the current study were offered as part of the regular teacher training curriculum of the module *Unterrichten*. In line with the research design, two cohorts were identified: the first cohort of prospective teachers that had just finished their practical semester and the second cohort of prospective teachers that had not yet participated in the practical semester. Each of the cohorts was randomly assigned to either a course designed according to the guidelines of the situated strategy or a course designed according to the guidelines of the cognitive strategy. To allow randomized allocation, the two courses related to a specific cohort were held concurrently.

The four courses were advertised regularly and identically in the online course-catalog of the university as “classroom management.” The description of the courses included working with video-cases but did not include specifics of either instructional strategies. Additional exposure to these courses was achieved by visiting all courses that the prospective teachers belonging to the population attended in the semester preceding the experiment: Participants of the first cohort were informed by visiting each of the eleven different courses accompanying the practical semester. Participants of the second cohort were informed by visiting each of the four different lectures of a mandatory introductory course. Participants that wanted to participate in one of these courses were asked to contact the course administrators. For each of the applicants, cohort-affiliation was verified using the university administrative system. Within each of the cohorts, the allocation of participants to the group was done randomly and before the pretest.

All courses took place in the summer of 2016 on three consecutive days. The courses related to the same cohort and were held concurrently. To support course-equivalence, the two cohorts

were separated one week from another.

6.3.2 Systematic similarities.

The development of both video-based courses followed several guidelines. In many respects, both video-based courses were identical. The following sections summarize all course-aspects that were the same for both courses. The consecutive sections summarize the course-aspects that were systematically varied. Figure 6. gives an overview of both video-based courses and the measuring instruments that were applied. Appendix C consists of the exercises in their original German wording.

Teaching objective

Both course-types had the same objective: to strengthen situation-specific skills in classroom management through the analysis of video-recorded cases. More specifically, both courses aimed at increasing prospective teachers' *perception*: the ability to filter out irrelevant information and focus on the relevant interactions and *interpretation*: to interpret the meaning and functions of these interactions and reconstruct the reasons for teacher behavior.

Classroom management strategies

Both course-types focused on increasing situation-specific skills in classroom management strategies. Chapter two showed that there is a broad variety of effective classroom management strategies. Following the guidelines regarding constructive alignment (Blomberg et al., 2013), the same classroom management strategies were selected for the video test CMC (see chapter 6.2) (Borich, 2011, 2014):

- Physical Classroom Arrangement (CA)
- Rules and Routines (RR)
- Low-Profile Management (LP)

Course-organization

The sessions of the courses were held on three consecutive days: Thursday, Friday, and Saturday. Each session lasted from 9:15 am to around 4:45 pm, excluding time for breaks and testing, in all

courses a total of 16 hours was reserved for learning activities. On each day, one of three classroom management strategies was addressed. The Thursday session focused on the analysis of the physical arrangement of the classroom, the Friday session focused on the analysis of rules and routines, and the Saturday session focused on the analysis of low-profile management. On Thursday, prospective teachers of both course-types attended the same lecture, in which the theme, relevance, and general approach to classroom management were introduced. Additionally, it was explained why the classroom setting typically leads to challenges relating to classroom order. After this introductory lecture, lasting around 60 minutes, all course-participants were directed towards their course room and were not brought together until the last section of the course, in which the post-test was administered. For each of the courses, a total of two course rooms were available.

A typical session had the following sequence: the day started with the teacher welcoming the course-participants and giving organizational instructions. After this, the morning learning block was held. After 60 minutes of lunch break, the afternoon learning block was held. The day was concluded with a short reflecting block, lasting around 30 minutes, in which that day's activities were summarized and reflected.

THURSDAY: Classroom Arrangement			FRIDAY: Rules and Routines			SATURDAY: Low-Profile		
Time	**Situated Strategy**	**Cognitive Strategy**	**Time**	**Situated Strategy**	**Cognitive Strategy**	**Time**	**Situated Strategy**	**Cognitive Strategy**
09:15 10:00	Introduction: Welcoming	Introduction Welcoming	09:15 09:45	Introduction Welcoming **FAIR-2-TEST**	Introduction Welcoming **FAIR-2-TEST**	09:15 11:40	Morning Session Welcoming E1: Explore Video 1: Which factors influence the success of the lesson? E2: Explore a hand-out: Which teaching aspects are coded using the presented coding-manual? E3: Coding of 20-second interval: Which teacher intervention was applied? (Video 1) E4: Coding of 20-second interval (Video 2) E5: Compare and evaluate	Morning Session Welcoming Short lecture on *Low Profile* E1: Code a 20-second interval: which teacher intervention was applied (Video 1)? E2: Coding of 20-second interval: which teacher intervention was applied (Video2)? E3: Comparing and evaluation
10:00 10:45	Lecture: Classroom Management The classroom as a social system and Classroom interaction		10:00 12:00	Morning Session E1: Explore contents of Video 1, Find overlap with yesterday's content, and overlap with role-model. Which teaching characteristics would you adopt? E2: Explore contents of Video 2, which characteristics of this teacher overlap with the role model, what would a coaching session between the teacher from V1 and V2 look like?	Morning Session Short lecture on *Rules and Routines* E1: Construct a chronicle of teacher and student agency. Apply observed rules and routines to the chronicle (Video 1) E2 Constructing a chronicle teacher and student agency. Apply observed rules and routines to the chronicle (Video 2) E3: Compare Results and evaluation			
11:00 13:00	Morning session E1: Define a teacher role-models on the basis of prior experience	Morning session Short Lecture on *Classroom Arrangement* E1: Map the Physical Arrangement of a classroom (Video 1) E2: Map the physical Arrangement of a classroom (Video 2)				11:40 12:00	**COURSE EVALUATION**	**COURSE EVALUATION**
13:00 14:00	*Lunch Break (60 min)*		12:00 13:00	*Lunch Break (60 Min)*		12:00 12:30	*Small lunch break (30 Min)*	
14:00 16:00	Afternoon Session E2: Formulate teaching challenges, alternatives of action (Video 1). E3: Formulate teaching challenges, alternatives of action (Video 2) E4: Compare both videos. Moderation at pin board	Afternoon Session E3: Identify Possibilities and barriers of Interaction (Video 1 and 2) E4: Compare, Evaluate and formulate alternatives Discussion of Results	13:00 15:30	Afternoon Session Group Discussion, the function of rules and routines E3: Advising a colleague, what could the teachers in V1 and V2 improve? E4: Compare, moderate and evaluation	Afternoon Session E4: Mark which routines were found in the video E5: Advising a colleague on the implementation of routines and rules E6: Comparing and evaluation	12:30 13:30	Afternoon Session E6: Compare results of Morning session with role-model	Afternoon Session E4: When to apply low-profile and high-profile intervention
16:00 16:45	Day closing	Day closing	15:30 16:15	Day closing	Day closing	13:30 14:00	Day closing	Day closing
						14:00 14:30	*Small break (30 Min)*	
						14:30 15:30	**POSTTEST CMC AND CME**	

Figure 6. **Overview of Exercises of Courses Following a Situated and a Cognitive Instructional Strategy**

Teachers

Two different teachers held the two concurrent courses. The first teacher was a senior professor in the pedagogy of secondary education, with several decades of experience teaching prospective teachers in the university. The second teacher was a PhD-researcher of pedagogy in secondary education with six semesters of experience teaching prospective teachers in the university. The senior professor started by teaching the course with the situated strategy. The PhD-researcher started out teaching the course with the cognitive strategy. To reduce teacher effects, both course-types were developed in agreement with both teachers. Additionally, both teachers switched groups after the second day of the course. For additional support concerning course-organization, for example for the distribution of testing and learning materials, in each group, a teaching-assistant was available for the whole period of the time.

Video clips

The number of video databases that are publicly available is limited (Petko et al., 2014). Especially videos depicting heavily challenged teachers in situations related to student discipline and lesson disturbances are often not publicly available. To find fitting video clips, four raters, three teachers and the author of the current study, searched through all available known video databases. For the selection process, the following criteria were applied. Firstly, to ensure ecological validity, the videos needed to depict teaching in classes of secondary education. Secondly, videos were selected to focus on only one of the classroom management strategies. The course designed according to the principles of the situated strategy would not receive theoretical input. Therefore, it was crucial that the content of the videos primarily related to only one of the classroom management strategies. Thirdly, the videos were selected to be relatively short, depicting one segment of classroom interaction. This was mainly done to focus learners' attention on one specific aspect of classroom management and, at the same time, by allowing a thorough analysis of that aspect through repeated watching. Fourthly, videos were selected based on their audio and visual quality. This criterion turned out to be challenging because the recording quality of some videos had been reduced for the sake of their use in online portals. Because course-participants would be able to watch the videos multiple times and use

headphones, it was decided that this criterion could be applied less strictly in favor of the first criterion.

After the selection process, seven videos were selected from two video corpora: the video corpus of Michael Hecht, recorded as part of the project called *Selbständigkeit im Unterricht: Empirische Untersuchungen in Deutschland und Kanada zur Paradoxie pädagogischen Handelns* (Hecht, 2009) and the video corpus of Professor Dr. Udo Rauin, recorded as part of the research project called *Strategien des Unterrichts in heterogenen Klassen und ihre Wirkung auf Schüleraktivität* (see J. Appel, 2016). For both course-types, these same seven video clips were used. These videos lasted between 2:36 and 6:00 minutes. The first strategy, the physical arrangement of the classroom, was supported by two video clips for the Thursday session. The second strategy, rules and routines, was supported by three video clips for the Friday session. The third strategy, low-profile management, was supported by two clips for the Saturday session. The video clips used for training purposes were different from the videos used in the video test CMC.

Learning arrangement and materials

Each course-participant was provided with a tablet computer for the duration of the session. The video clips were published on an online video-based learning platform called VIGOR (Level: Lehrerbildung vernetzt entwickeln, 2016). Course-participants connected their tablets to VIGOR and watched the video clips on their tablet computers with headphones. During the sessions, course-participants could watch the video clips as often as they wanted, with no restrictions on replaying, pausing, and skipping to a relevant part of the video clip. Moreover, each of the course-participants was provided with a hand-out, consisting of all exercises of that day. To standardize the courses, the same hand-outs were used for both cohorts. Course-participants typically worked in groups of four to five participants.

Certificates and grading

Courses within the curriculum of Bildungswissenschaften are concluded with a certificate of participation or a graded essay, in which one aspect of classroom management is reflected on the basis of scientific literature and the video-analysis in the course. All students were given a

choice to finish the course with either the certificate of participation or the graded course paper. The issue of certificates was dependent on regular attendance in the courses and active participation in the course. The grades of the essays were only dependent on the quality of scientific reflection. The activities within the courses and the scores on the video tests did not influence the grades.

6.3.3 Systematic differences.

The courses were systematically varied on the principles of the situated and cognitive instructional strategy as discussed in chapter 4.3 and oriented on the studies by Tina Seidel and Geraldine Blomberg (Blomberg et al., 2013; Blomberg et al., 2014; Seidel et al., 2013). The systematic differences center around five different teaching aspects: (1) the function of video clips, (2) the function of theory, (3) the role of the teacher and instruction, (4) the degree of structure within the exercises, and (5) the function of previous practical experience.

Function of video clips

In both courses, videos had a different function. In the course following a cognitive strategy, videos are used to apply previously conveyed theoretical knowledge in a systematically structured method. It was assumed that situation-specific skills in classroom management are acquired through the strict application of an analytical method so that cognitive load produced by working with videos is reduced (Van Merriënboer & Kirschner, 2009). Consequently, videos were introduced after a theoretical introduction to the topic and a methodological introduction to the way the video would be analyzed. Alternatively, in the course following a situated strategy, videos were used as an anchor to activate previous experiences from which regularities could be derived that would lead to the formulation of general valid rules. It was assumed that situation-specific skills in classroom management are acquired by activating previously acquired practical experiences and connecting these experiences to the cases presented in the videos (Korthagen, 2010). Consequently, videos were introduced without providing theoretical or methodological input.

Function of theory

Theory plays a different role in both course-types. In the course following a cognitive strategy, theoretical knowledge is conveyed before the analysis of the video clips. At the start of each session, the classroom management strategy of that day was introduced by the teacher in a lecture lasting between 30 and 45 minutes. In this lecture, the relevance and several examples of the classroom management strategy were provided for the whole group. In the course-type designed after the guidelines of the situated instructional strategy, no theoretical input was given before the start of the video analysis. Instead, course-participants were asked to formulate effective and less-effective teacher behavior as seen in the video clips. From this, course-participants were asked to formulate the corresponding general applicable rule or classroom management strategy.

Teacher instruction

In the course following the cognitive strategy, before the group-work, the teacher presented theory and provided direct instruction on how the video-cases should be analyzed. During the group-work, the teacher helped each of the groups with clearing questions regarding the applied method. After the group-work, the teacher summarized, discussed, and evaluated the work of the groups in the plenum. In the course following the situated strategy, the teacher did not provide direct instruction regarding the theory and method of video-analysis. Instead, the teacher only provided instruction on organizational aspects like the available time and learning materials. During the group-work, the emphasis of the teacher was on stimulating discussion within the groups and the stimulation of recalling previous practical experiences. After the group-work, the teacher took over a moderating role, collecting, arranging and discussing the findings of the groups. This moderation phase aimed to formulate general rules or classroom management strategies.

For the Thursday session on rules and routines, the teacher-instruction slightly deviated from this principle. As it was evaluated that the content of the available video clips itself did not provide adequate cause for the prospective learners to discuss the topic of routines, it was decided that the teacher introduced this concept also in the course following a situated instructional strategy.

To reduce the impact of this deviation, the teacher introduced the concept only at the start of the afternoon session and in the format of a group conversation instead of a lecture.

Video-analysis exercises

Cognitive load theory assumes that the initial complexity provided by video-cases should be reduced by breaking down the analysis in individual highly structured steps (see chapter 4.3.3). Consequently, the exercises within the cognitive course consisted of a step-by-step plan aimed at the systematic analysis of the video-cases. For example, in one of the exercises of the Thursday session, related to the physical arrangement of the classroom, course-participants drew a map of the classroom in which they marked the viewing angles of all individual students and the teachers, the physical movement of the teachers, and visually marked resulting barriers. Alternatively, in the situated course, participants were asked to explicate their personal practical experiences and connect these to the situation seen in the video clip. Consequently, the exercises within the situated course focused initially on the discussion of course-participants' personal experiences in school. No method was provided for the analysis of the video-cases. The exercises of the Thursday session on the arrangement and the use of the physical arrangement of the classroom asked participants to view the video-analysis with the question what they found striking, which teacher behavior they thought was effective for classroom order and which teacher behavior was thought to be ineffective and to compare these to their own practical experiences in school settings.

An exception to this principle was made for the Saturday session on low-profile. Similar to the videos of the Thursday session regarding rules and routines, it was expected that the depicted classroom interaction in the video clips of that day did not depict situations that would suffice for the course-participants to initiate a discussion on the topic of low profile on their own accord. Therefore, the exercises of both seminars involved a systematic analysis of low-profile teacher-interventions.

Function of practical experience

Corresponding to the guidelines of situated learning (Korthagen, 2010), participants of the

courses with a situated instructional design were asked on multiple occasions to bring up their own experiences as part of their exercises. Considering the setup of the research design, these experiences differed between the two cohorts. Therefore, the instructions were formulated so that they could relate to the practical experiences they had as a student as well as the practical experiences they collected during the practical semester. Alternatively, in the courses with a cognitive instructional strategy, course-participants were not explicitly asked to bring up memories of their own experience in schools.

6.4 Instruments

Apart from the new video test CMC, data from other instruments were collected. In the following sections, these instruments are discussed in more detail. The video test CME (König, 2015b) (6.4.1), FAIR-2 (Moosbrugger & Oehlschlägel, 2011) measuring attentiveness (6.4.2), items measuring self-efficacy on classroom management (Pfitzner-Eden et al., 2014) and general cognitive achievement (6.4.3), and items evaluating course-quality (6.4.4).

6.4.1 Classroom management expertise (CME).

To assess convergent validity, the validity-type related to construct validity (see chapter 6.1.2), the video test CME (König, 2015b) was assessed in parallel to the new video test CMC. Although both tests measure two different alternatives of situation-specific skills on classroom management, it is assumed that both constructs are related and therefore parallel measurement assesses convergent validity.

CME was administered directly after CMC in both pretest and posttest. CME was administered using the same procedure as CMC: test-participants watched the video clips on a centrally placed screen and processed the questionnaire on a tablet computer with a wireless keyboard. Data of CME and CMC were gathered in real-time by the locally installed server running the software LimeSurvey (LimeSurvey, 2016).

Coding of CME was done using the official coding manual and using the technique described in chapter 6.2.6: codes were entered directly into a specially-formatted spreadsheet. Four raters were trained to code the data using the instructions from the coding manual of CME. The four

raters were divided into two pairs. Each pair of raters independently coded half of the data in parallel. Because of illness, one rater needed to be replaced by a fifth rater. The new rater received training in the use of the CME coding manual and took over the coding of the original rater.

6.4.2 Attentiveness: FAIR-2.

FAIR-2 is a standardized and validated instrument for the measurement of attentiveness (Moosbrugger & Oehlschlägel, 2011; Petermann, 2011). FAIR-2 consists of a paper-pencil questionnaire with multiple rows of symbols. Test-participants are asked to evaluate these symbols row by row by drawing a line with a pen or pencil underneath these symbols. Whenever a symbol meets the criteria, participants are asked to mark the symbol by moving their pen upwards. Figure 7. shows an example of a row of symbols where a participant was asked to mark the symbols that consist of a square with two dots and a circle with three dots.

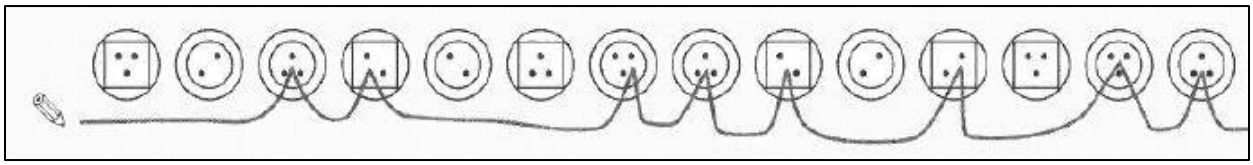


Figure 7. **Item-examples of FAIR-2 (Moosbrugger & Oehlschlägel, 2011, p. 34)**

FAIR-2 can be considered a speed-test because there is a six-minute time-limit and most participants will not be able to evaluate all the symbols within this time. The test assumed that a person's attentiveness is expressed in three ways. Each of these kinds of attentiveness is represented by a separate score. Firstly, attentiveness shows in the number of symbols that are correctly marked within the available amount of time. People with a high degree of attentiveness are expected to process many items in the available time. This ability is expressed by the *performance-score* (P) that represents the number of successfully marked symbols within the time-limit. Secondly, a person's attentiveness is shown by a low number of mistakenly evaluated symbols. A mistake includes (a) marking a symbol that does not meet the criteria and (b) skipping a symbol that does meet the criteria. This kind of attentiveness is expressed by the *quality-score* (Q) that expresses the ratio of successfully marked symbols to the total number of processed

symbols. Thirdly, a person's attentiveness shows by the degree to which a test-participant can maintain attentiveness throughout the available time. When a participant has worked very fast, he or she might have a relatively high P-score, but also have a higher proportion of mistakes, resulting in a relatively low Q-score. Contrary, a test-participant working slowly might have a relatively low P-score, but therefore hardy make mistakes resulting in a high Q-score. The *continuity-score* (C) represents both aspects of attentiveness: the performance and the quality and is calculated by the multiplication of P- and Q-scores. FAIR-2 was administered on the morning of the second day of the course.

FAIR-2 was administered using paper-pencil questionnaires and coded using the official templates. The codes were then further analyzed using the official software (Moosbrugger, Oehlschlägel, & Steinwascher, 2017).

6.4.3 Scales from the practical semester: self-efficacy and cognitive achievement.

Either before or after partaking in the experiment-related courses, all prospective teachers of the study participated in the practical semester (see Figure 1.). The practical semester was evaluated by an extensive research project initiated by the Hesse government (see chapter 6.3.1). Several of the scales that were used as part of this research project offered relevant data. All participants were asked to evaluate their classroom management self-efficacy at the start of their internship on a 6-point Likert-scale (1 = not at all certain, 6 = completely certain (Pfitzner-Eden et al., 2014). Figure 8. lists the items that were used.[6]

Item
How certain are you that you can...
...control disruptive behavior in the classroom?
...get students to follow classroom rules?
...calm a student who is disruptive and noisy
...keep a few problem students from ruining an entire lesson

Figure 8. **Items of Perceived Self-Efficacy in Classroom Management (Pfitzner-Eden et al., 2014, p. 88)**

[6] English formulations of the items were taken from Pfitzner-Eden et al. (2014, p. 88).

Chapter 3.6 discusses the construct of teacher's perceived self-efficacy. Here, it is stated that teacher self-efficacy is a related and relevant aspect of teachers' competence. However, because it does not represent an objective measurement, it should not be considered equivalent to more objective competence aspects as discussed in the model of situation-specific skills. Therefore, the first purpose of the scores on self-efficacy on classroom management is to calculate discriminant validity. Considering the nature of both aspects of competence, no significant correlation between CMC and scores of classroom management self-efficacy was expected.

Additionally, the scores on self-efficacy were used to investigate how participation in the video-based courses influenced prospective teachers' perceived self-efficacy at the start of the practical semester. As discussed in chapter 3.6, studies report a positive effect of video-based courses on prospective teachers' self-efficacy in classroom management within a pre-post research design (Gold et al., 2017; Kumschick et al., 2017). As part of the evaluation program of the practical semester (Practical Semester Hesse, 2018), the instrument of perceived self-efficacy on classroom management was applied multiple times during the practical semester. However, because of relatively high degrees of attrition, only the data from the first measurement point, at the start of the practical semester, were used. Prospective teachers of the first cohort participated in the video-based courses after the practical semester. Prospective teachers of the second cohort participated in the video-based courses before the practical semester. To measure the effect of participation in a video-based course on self-efficacy on classroom management, the values of self-efficacy of the two cohorts were compared.

As part of the research project of the practical semester, participants were also asked to state two measures of general cognitive ability: the grade they received in the introduction course of *Bildungswissenschaften*, called BW-A and the average grade of their high-school diploma. Both grades are used for the measurement of discriminant validity of CMC, as it is assumed that the construct measured with the video test is independent of general cognitive ability as expressed by the two grades.

6.4.4 Course-evaluation.

Goethe University Frankfurt uses a standardized questionnaire for the evaluation and quality-

management of all courses (LUQ, 2018) (See Figure 10.). The questionnaire consists of items regarding different course aspects like the perceived gains in knowledge, course-structure, use of learning materials, and the appropriateness of the lecture room with a 6-point Likert-scale. Values on these items were used to gain insight into the subjective experience of the course participants. It was firstly investigated to what degree these subjective experiences differed between the two cohorts. It was assumed that prospective teachers from the first cohort due to their participation in the practical semester were more acquainted with classroom interaction than prospective teachers from the second cohort (see chapter 4.2). Therefore, the question arises, if differences in the experienced difficulty level and pace of the course existed. To answer this question, values on items 1. and 2. from the questionnaire (see Figure 9) were used. Secondly, it was investigated, to what degree these subjective experiences differ between the two course-types. As the systematical difference between the two course-types relates to the degree of structure and the didactical integration of the video clips, the question arises, to what degree these differences in course-design were also experienced by the participating prospective teachers. To answer this question, values on items 3 and 4 were used (see Figure 9.). As the other questions from the questionnaire were not directly related to the above-listed questions, their data were not analyzed. The questionnaires were administered on the last day of the course, directly after finishing the last learning block of the course and before the posttest of both video tests.

Item
1. Course-participation leads to a noticeable increase in knowledge.
2. The learning pace of the course is appropriate.
3. The subject matter of the course is well structured.
4. Materials and media are used appropriately in the course.

Figure 9. **Items on Course-Evaluation**[7]

[7] German to English translations made by the author.

6.5 Sample

6.5.1 Population.

The target population consists of two cohorts of prospective teachers of the higher secondary education teacher training track of Goethe University Frankfurt. The first cohort consists of prospective teachers who have participated in the practical semester. As the practical semester typically takes place in the third term, the population consists of all prospective teachers enrolled in their fourth term. In the summer semester of 2016, the semester in which the intervention took place, a total of 237 prospective teachers were registered in their fourth term, of which 141 (59.5%) were female. The second cohort consists of prospective teachers who have not yet participated in the practical semester. The population of this cohort consist of all prospective teachers enrolled in their second term. In the summer semester of 2016, a total of 265 prospective teachers were registered in their second term, of which 152 (57.4%) were female (Goethe University Frankfurt). The age of the population varies, ranging from 19 to even over 40, because the programs of teacher training of Goethe University are characterized by a relatively high percentage of lateral entry students.

6.5.2 Registration and selection process.

Prospective teachers have a high degree of freedom to choose the courses in which they want to participate. Before the start of the semester, a total of 120 participants had expressed their interest in participating in the courses. After checking that all participants met the conditions of enrollment in the targeted program of teacher training and belonging to the right cohort - either with or without experience in the practical semester, 115 prospective teachers were registered for the courses. After the preparatory sessions were held, seven participants canceled their registrations. Eight other participants did not show up for the pretest, leaving a total of 101 prospective teachers to participate in the pretest. Of this group, 50 belonged to cohort 1 and 51 belonged to cohort 2. Of these 101 prospective teachers in the pretest, 13 could not be included in data-analysis because they either dropped out after pretest or their pseudonyms of pretest and posttest could not be matched. The total sample that was included in data-analysis consisted of n = 88 participants. Because courses on classroom management are not mandatory, the

research sample is considered a convenience and self-selected sample and can, therefore, not be considered representative of the population (Döring & Bortz, 2016b).

6.5.3 Sample.

Recent studies indicate that more than 50% of the prospective teachers gain additional teaching experience through part-time jobs, for example, as a substitute teacher (Bäuerlein & Reintjes, 2018). To investigate the degree to which course-participants had gained teaching experience outside of the practical semester, two items were added to the pretest: “Did you exert any teaching activities in school, for example, as a substitute teacher?” “If yes, for how many months?” Analysis of frequency distribution showed that 46 participants answered ‘no’ to the first question. The 42 participants that answered ‘yes’ to the first question were strongly spread out with a minimum value of 1 and a maximum of 24. Values were recoded into three groups to get an evener distribution: participants having no teaching experience outside of the practical semester, participants having between 1 and 6 months of teaching experience and participants having seven or more months of teaching experience. The frequency distribution for each of the groups is reported in Table 7.

The sample of the first cohort consisted of n = 44 participants who had previously partaken in the practical semester. Of these 44 participants, 56.8% were female, and 43.2% were male. With one participant not stating his age, participants’ age ranged from 19 to 28 (M = 21.9; SD = 2.3). The sample of the second cohort also consisted of n = 44 participants, who had not yet partaken in the practical semester. Of these 44 participants, 63.6% were female and 36.4% male. With one participant not stating his age, participants’ age ranged from 18 to 37 (M = 22.1; SD = 4.5). Randomized allocation of the participants to the groups took place before the pretest. As some sample attrition occurred after the allocation, the groups were not evenly sized. Table 7 gives an overview of the sample size of the different groups and their respective values for age and gender. Table 7 shows that no big differences existed regarding age and gender. However, it can be seen that larger differences between the two cohorts existed regarding the number of months of teaching experience. Add to this the fact that the first cohort had participated in the practical semester; the two cohorts were considered equivalent.

Table 7
Sample Size, Gender, Spread, Mean and Standard Deviation of Age, and Teaching Experience for all Four Groups

Group	Sample Size	Gender (% Female)	Age			Distribution Experience		
			min. – max.	*M*	*SD*	0 months	1 - 6 months	7 - 24 months
1Sit.	24	54.2	19 – 28	22.1	2.34	5	14	5
1Cog.	19	60.0	20 – 28	21.7	2.36	6	9	5
2Sit.	19	60.0	19 – 30	22.2	3.27	16	2	2
2Cog.	24	66.7	18 – 37	22.0	5.37	19	3	2

Note. 1Sit. = first cohort, situated strategy, 1Cog. = first cohort, cognitive strategy, 2Sit. = second cohort, situated strategy, 2Cog. = second cohort, cognitive strategy.

6.6 Research Ethics

Before the start of the experiment, participants of each of the cohorts visited a preparatory session. In compliance with ethical directives (Döring & Bortz, 2016a), all participants were informed that the courses were part of a research project regarding classroom management competence of prospective teachers. No information was given about the systematic differences between both courses. Participants were also informed about the assessment procedures of the research project and that participation in the research project, video tests and courses were voluntary, that data was acquired anonymously, and test-scores did not affect the grading of the course. Additionally, all participants signed a declaration of discretion, expressing that they would not copy or otherwise digitally extract the videos used in the course and test and that their contents would not be discussed in with third parties. No information was given on the experimental conditions and hypothesis. Participants of the second cohort attended the preparatory meeting on 19 April 2016. Participants of the first cohort attended the preparatory meeting on 26 April 2016. Before the start of each assessment, participants were reminded of voluntary participation and what participants could skip a question if they did not want to answer. Participants were urged to write down the answer that seems the most likely, but that they could skip the question if they did not know the answer.

6.7 Hypotheses for statistical data-analysis

In chapter 5, the main research question was formulated as follows:

How can classroom management competencies be effectively conveyed in video-based courses of teacher training programs with regard to the differences in practical experiences of prospective teachers?

Following the methodology described in this chapter, the hypotheses are divided into hypotheses related to the construct validity of the video test CMC and hypotheses related to the evaluation of the effectiveness of the courses. The hypotheses regarding construct validity of the CMC video test are grouped in hypotheses regarding convergent validity and hypotheses regarding discriminant validity (Döring & Bortz, 2016b) (see chapter 6.1). Convergent validity refers to the positive correlation of test scores of CMC with values that are theoretically related. Because of the similar test-structure and the measured construct, it was expected CME and CMC significantly correlate at both pretest and posttest (H1a). Regarding discriminant validity, it was expected that scores on CMC did not correlate to attentiveness, as expressed by P-, Q-, and C-Scores of the FAIR-2 test (H1b). Because situation-specific skills and self-efficacy are regarded as different aspects of competence, scores on self-efficacy regarding lesson disturbances at the beginning of the practical semester were not expected to correlate to scores on CMC (H1c). Because the development of self-efficacy is assumed to be dependent on mastery experiences, vicarious experiences and social persuasion (see chapter 3.6), the hypothesis of discriminant validity was only measured for the cohort that scored their self-efficacy on classroom management after they participated in the video-based courses. Because situation-specific skills are not considered related to general cognitive performance, scores of general achievement like the grades of the high-school diploma and the grade of the introductory BW-A-Module were not expected to correlate (H1d).

The hypotheses regarding the evaluation of the effectiveness of the courses consist of one hypothesis regarding self-efficacy and several hypotheses regarding situation-specific skills. As one study showed the effectiveness of video-based courses on perceived self-efficacy of prospective teachers (see chapter 3.6), it was expected that prospective teachers from the first

cohort whose self-efficacy was measured before participating in the video-based courses was lower than the prospective teachers from the second cohort, whose self-efficacy was measured after participation in the video-based courses (H2a). Regarding situation-specific skills on classroom management, firstly, a strong main effect was expected, showing that on average, all participants increased their scores on CMC (H2b). Secondly, no significant differences were expected regarding the interaction effects of the first order. No differences in the change in CMC-scores exist between the two cohorts were expected (H2c) and no difference in change between the two course-types was expected (H2d). Thirdly, regarding second-order interaction effects, it was expected that for the cohort that partook in the practical semester, CMC-scores of participants in the course with a situated strategy increase significantly more than the CMC-scores of participants in the course with a cognitive strategy (H2e). For the cohort that did not partake in the practical semester, CMC-scores of participants in the course with a cognitive strategy increase significantly more than the CMC-scores of participants in the course with a situated strategy (H2f).

6.8 Data-Analysis

The data of the different measurements (pretest, posttest, FAIR-2, evaluation of the practical semester, and course-evaluations) were matched using pseudonyms. On each measurement, test-participants were asked to construct a pseudonym based on the first two letters of their mother's first name, the first two letters of their father's first name, the first two letters of the first name of the grandmother of their mother side, the first letter of the town of birth and the day of their birth. In some cases, pseudonyms could not be matched, leading to missing values. The number of failed matches of a pseudonym is discussed in more detail in chapter 7.3.

The hypotheses formulated in chapter 6.7 were tested using the software SPSS 23. The hypotheses regarding convergent (H1a) and discriminant (H1b, H1c, and H1d) validity were tested by calculating correlation coefficients between scores of CMC and scores of one of the tests described in chapter 6.4. Because of the sample size, a normally distributed sampling distribution can be assumed as a result of the central limit theory. (Field, 2013). Hypotheses on test-validity were therefore tested calculating Pearson's r coefficient in SPSS 23. The hypothesis

regarding the effects of the treatment effects on self-efficacy (H2a) was tested by calculating the independent-sample t-test in SPSS 23. The hypotheses regarding the effects of the treatment on situation-specific skills (H2b to H2f) were tested calculating a mixed-design ANOVA that combines a two-way ANOVA with a repeated-measurement ANOVA using SPSS 23 (Field, 2013).

7. Results

This chapter reports the empirical findings related to the main questions of the current study. The first three sections report the results of an investigation of the newly developed video test CMC. Firstly, the results of inter-rater reliability for both video tests: CME (König, 2015b) and the newly developed video test CMC are reported (7.1). Secondly, final revisions to the video test based on item-analysis and analysis of internal consistency are reported (7.2). Thirdly, the factorial structure of CMC was investigated through exploratory factor analysis (7.3). Hence, the subsequent section reports the results of descriptive statistics for each of the applied measurement instruments (7.4). For each of the instruments, missing values, mean values, dispersion, and outliers are reported per group and assessment. The results of a validation study are reported through testing the hypotheses regarding the construct validity (7.5). The last section of this chapter reports the results of the tests regarding the hypotheses about the effectiveness of the video-based courses (7.6).

7.1 Inter-Rater Reliability

Inter-Rater reliability or *concordance* refers to the degree of agreement among different raters (Wirtz, 2002). The values of the index of inter-rater reliability indicate the degree of scoring objectivity (Ch 6.1.2.1) (Döring & Bortz, 2016b). High values on an index for inter-rater reliability shows that there was a small effect of the rater and his or her subjective interpretation of coding instructions on the score of a specific criterion. Low values on an index for inter-rater reliability indicates differences in the interpretations of the coding instructions.

A typical index with which inter-rater-reliability (IRR) is reported is percent agreement (p_o) which is calculated by dividing the number of scores where two raters agree by the total amount of codes that were given. However, p_o is not corrected for the agreement that can occur by chance alone: with dichotomous variables, for example, there is a 50% chance of a random agreement. To resolve this issue, the chance-corrected index of Cohen's Kappa (κ) is often reported. However, the disadvantage of Cohen's Kappa coefficient lies in the fact that it is less useful for variables that are strongly positively or negatively skewed (Wirtz, 2002). For the calculation of

IRR on dichotomous variables with high or low values of item-difficulty, the paradoxical situation can occur that despite very high levels of percent agreement, values for Kappa are very low or even negative (Feinstein & Cicchetti, 1990). To solve this issue, two additional indices of IRR are used: p_{pos} and p_{neg}. These indices indicate the percentage of the cases, in which the two raters both pass a positive (or negative) judgment when one of the raters has already passed a positive (or negative) judgment (Wirtz, 2002, p. 90) and are less prone to skewed distributions. Because no omnibus index exists that solves all paradoxes for dichotomous variables, it is typically advised to report all three important indices (Cicchetti & Feinstein, 1990; Wirtz, 2002). Indices for IRR were calculated using the software Winpepi (Abramson, 2016).

According to general rules of thumb, the reported average values of indices of IRR can be regarded good for CME ($.60 < \kappa < .75$) and very good for CMC ($\kappa > .75$) (Wirtz, 2002, p. 59). The values for variability indicate that despite the good values for IRR for CME and CMC, there are several criteria for which values of IRR-indices were lower. For both video tests, the raters discussed the criteria where the codes initially differed and decided on the final score that was used for further analysis.

Table 8
Indexes of Inter-Rater Reliability for CMC

IRR-Index	*M*	*SD*	*Min.*	*Max.*
Cohens κ	.77	.18	.15	.99
p_o	92.9	5.6	77.1	99.5
p_{pos}	91.1	9.2	59.3	99.6
p_{neg}	85.6	15.6	21.7	99.2

Note. κ *= Cohens Kappa,* p_o = percentage agreement, p_{pos} = positive agreement, p_{neg} = negative agreement

Table 9
Indexes of Inter-Rater Reliability for CME

IRR-Index	*M*	*SD*	*Min.*	*Max.*
Cohens *κ*	.73	.21	.04	.97
p_o	91.4	7.34	70.7	99.0
p_{pos}	91.8	9.06	54.9	99.4
p_{neg}	80.4	19.0	8.3	99.1

Note. κ = Cohens Kappa, p_o = percentage agreement, p_{pos} = positive agreement, p_{neg} = negative agreement

7.2 CMC: Missing values, Item-Analysis, and Final Revisions

Before carrying out the data-analysis regarding the main question, several final changes to CMC were made. Missing values were imputed using the method of multiple imputation (7.2.1). Based on the results of item-analysis and analysis of internal consistency of the data of the main study (7.2.1), several items were eliminated (7.2.2). After completion of this process, a final version of CMC consisting of 20 items was constructed of which the results of item-analysis and internal-consistency are reported (7.2.2).

7.2.1 Missing values.

For the calculation of missing values in CMC, the guidelines of the CME video test (König, 2015b) were applied. According to the coding guidelines of CME, items should be coded as missing when test-participants did not complete all questions in the available time. Analysis of missing values showed that at pretest 1.0% and at posttest 0.7% of all values were missing. At pretest, missing values occurred on 5 of the 30 items, at posttest on 4 of the 30 items. At pretest, 24 of 101 participants (23.8%) showed missing values. At posttest, 17 of 90 test-participants (18.9%) showed missing values. Although the total number of missing values can be considered low, they concern over 20% of the test-participants. These test-participants would be left out from further analysis when missing values were left untreated (Eid, Gollwitzer, & Schmitt, 2015). To be able to include the cases that showed missing values into further analysis, it was decided to impute missing values (Döring & Bortz, 2016b).

The decision to use the method of imputation is based on the reason why missing values occur.

Missing values can occur *completely at random* (MCAR). Here, the fact a value is missing is unrelated to the underlying or another relevant variable of the test-participant. Missing values can also occur *at random* (MAR). Here, the fact that a value is missing is related to relevant variables, but the value of these related variables is known and therefore the rate to which an item is not answered can be predicted. Lastly, missing values can occur systematically, or *not at random* (MNAR). Here, missing values are related to relevant variables, but the values of these variables are unknown, making it impossible to predict the values for unanswered items (Eid et al., 2015). A systematic bias in missing values would relate to the incompetence of the test-taker. In this case, the number of missing values would significantly correlate with the mean score of the other items. Analysis showed that there was no significant correlation between the number of missing values and mean item-score for pretest, τ = -.053, p = .52, and posttest, τ = -.070, p = .42. These results indicate that missing values occurred completely at random (MCAR). Literature advises the method of multiple imputation in which for each missing value, multiple values are estimated based on the values on the other variables (Eid et al., 2015).

For each of the missing values, five imputations were calculated using the software SPSS 23. Typically, further data-analysis is carried-out on the pooled data of these five imputations. However, the method of repeated measures analysis of variance does not allow the use of pooled data. To create a data-matrix that can be used for further analysis, the median for all five imputations was calculated and fed back into the data matrix.

7.2.2 Item-analysis.

As described in chapter 6.2, the video test named CMC was developed to measure situation-specific skills on classroom management. A first draft of the test was piloted, after which items were evaluated and reformulated. Model-answers were formulated over the course of two rounds of expert-ratings. The relative weight of each of the model answers was defined through a third round of expert-ratings. The coding rubric, in which each of the model answers was formulated into a criterion, was developed inductively over the course of multiple rounds of pilot coding. For the assessment in the main study, CMC consisted of 30 Items, of which 22 were open-response, six items were single-choice, and two items consisted of 5 or 6 true-false statements.

The open-response and true-false items were designed to be trichotomous, consisting of item-scores of 0, 1, and 2. The single-choice items were designed to be dichotomous, consisting of item-scores of 0 and 2. Items were divided relatively evenly across the three categories of each theoretical dimension (see Table 5). The calculation of the item-scores was based on the rating of the third round of expert-ratings and an item-analysis on the criterion-level. These item-scores were then used for further analysis.

Table 10 gives an overview of the results of an item-analysis of these item-scores. It shows the result of all participants in pretest (n = 101) and posttest (n = 90). Chapter 6.1 discusses the main criteria for the evaluation of item and test-quality. Below, these criteria are only summarized. The item-difficulty coefficient (*P*) is calculated by dividing the average item-score divided by the maximum score for that item (Moosbrugger, 2012b). Low values indicate that an item was hard for the test-participants to answer. High values indicate that an item was easy for the test-participants to answer. Together with the distribution of scores, this coefficient indicates to what degree the items are evenly divided across the difficulty spectrum. Analysis of the item-difficulty-coefficient shows that at pretest, the mean item difficulty lies at $P = .54$ ($SD = .15$), 27 items' values fall within the range of $.25 < P < .75$, one item has a value below .25 and two items have a value above .75. At posttest, mean item-difficulty lies at $P = .58$ ($SD = .18$), 23 items' values fall within the range of $.25 < P < .75$. One item has a value below $P < .25$, and six items have a value above $P > .75$. Guidelines for test-development recommend including items that cover the whole spectrum of item-difficulty and that mostly fall within the middle range (Bühner, 2011). The results of item-analysis show that this is the case for the 30 items of CMC, indicating that CMC's difficulty is appropriate for measurement of the sample.

Table 10

Item-Analysis of CMC-30

Item			Descriptives Pretest						Descriptives Posttest						Reliability Pretest		Reliability Posttest	
			Score-Distribution						Score-Distribution									
ID	Sit. Spec. Skill	CM-Dimension	*M*	*P*	0	1	2	Var(x_i)	*M*	*P*	0	1	2	Var(x_i)	r_{it}	Alpha if Item Deleted	r_{it}	Alpha if Item Deleted
1.1	AP	LP	0.99	.50	10	82	9	.19	1.14	.57	7	63	20	.28	.19	.49	.40	.32
1.2	AP	CA	1.15	.57	43	x	58	.99	0.42	.21	71	x	19	.67	-.08	.53	-.07	.40
1.3	JA	CA	1.12	.56	6	77	18	.23	1.36	.68	2	54	34	.2778	.20	.49	.32	.34
1.4	HP	LP	1.15	.57	43	x	58	.99	0.91	.46	49	x	41	1.00	-.00	.51	-.01	.39
1.5	HP	LP	0.93	.47	41	26	34	.75	0.94	.47	40	15	35	.84	.15	.49	-.18	.43
1.6	HP	LP	0.58	.29	52	39	10	.45	0.63	.32	42	39	9	.44	.26	.48	.13	.36
1.7	AP	LP	1.01	.50	50	x	51	1.01	0.84	.42	52	x	38	.99	.01	.51	-.13	.42
1.8	JA	RR	1.13	.56	36	16	49	.83	1.34	.67	19	21	50	.66	.23	.47	.00	.38
2.1	AP	CA	0.91	.46	30	50	21	.50	1.07	.53	16	52	22	.42	.08	.50	.07	.37
2.2	JA	LP	0.78	.39	42	39	20	.57	1.14	.57	22	33	35	.62	.16	.49	.24	.34
2.3	AP	CA	1.03	.51	30	38	33	.63	1.40	.70	13	28	49	.54	.32	.46	.06	.37
2.4	AP	LP	1.02	.51	14	71	16	.30	1.12	.56	10	59	21	.33	.22	.48	.09	.37
2.5	AP	RR	0.75	.38	63	x	38	.95	1.09	.54	41	x	49	1.00	.01	.51	.08	.37
2.6	JA	LP	1.68	.84	9	14	78	.40	1.79	.89	5	9	76	.28	.40	.46	.23	.35
2.7	HP	RR	1.35	.67	33	x	68	.89	1.53	.77	21	x	69	.72	.16	.48	-.01	.39
3.1	AP	LP	1.47	.73	14	26	61	.53	1.49	.74	8	30	52	.43	.07	.50	.15	.36
3.2	JA	LP	1.18	.59	36	11	54	.87	1.58	.79	15	8	67	.58	-.04	.52	-.04	.39
3.3	HP	CA	1.35	.67	27	12	62	.77	1.62	.81	14	6	70	.55	.23	.48	.10	.36
3.4	AP	RR	1.29	.64	36	x	65	.93	1.29	.64	32	x	58	.93	.15	.49	-.05	.40
3.5	JA	RR	1.09	.54	37	17	46	.82	1.03	.52	39	9	42	.91	.09	.50	.21	.34
3.6	AP	RR	1.03	.51	30	38	33	.63	1.34	.67	8	43	39	.41	.15	.49	.06	.37
3.7	HP	RR	1.18	.59	29	25	47	.73	1.31	.66	22	18	50	.71	.19	.48	.30	.32
3.8	JA	CA	1.24	.62	18	41	42	.54	1.06	.53	26	33	31	.64	.11	.49	.07	.37
4.1	AP	RR	0.79	.40	21	80	0	.17	0.70	.35	33	51	6	.35	.18	.49	.01	.38
4.2	AP	LP	1.38	.69	19	25	57	.62	1.60	.80	10	16	64	.47	.20	.48	.18	.35
4.3	HP	RR	1.39	.69	20	22	59	.64	1.64	.82	8	16	66	.41	.16	.49	.43	.31
4.4	JA	RR	0.70	.35	54	23	24	.69	0.84	.42	46	12	32	.85	.21	.48	.12	.36
4.5	AP	CA	0.59	.30	71	x	30	.84	0.78	.39	55	x	35	.96	-.02	.51	.17	.35
4.6	JA	CA	1.53	.77	15	17	69	.55	1.59	.79	14	9	67	.56	.20	.48	.12	.36
4.7	HP	LP	0.34	.17	84	x	17	.57	0.36	.18	74	x	16	.59	.05	.50	-.02	.39

Note. AP = accuracy of perception; HP = holistic perception; JA = justification of action; CA = classroom arrangement; RR = rules and routines; LP = low profile; M = Mean; P = item-difficulty; Var(x_i) = item-variance; r_{it} = item-total correlation,

Item-variance Var(x_i) refers to the degree that item-scores vary across the sample. As such, it is an indicator of the item's ability to differentiate between test-participants with different levels of the variable (Moosbrugger, 2012b). As a target, item-variance should be high, because this increases its ability to differentiate (Moosbrugger, 2012b). Analysis of item-variance shows that at pretest, mean item-variance lies at M = 0.65 (SD = 0.24) with 28 items having a value above Var(x_i) > .20. At posttest, mean item-variance lies at M = 0.61 (SD = 0.23), with all 30 items having a value above Var(x_i) > 0.20. The six single-choice fixed-response items and two open-response items were designed in such a way that they either gave the full score (2) or no score (0). Because the middle score (1) was not available, these items have relatively high scores on item-variance and increase the mean values of item-variance for all 30 items. Despite this bias, regarding item-variance of the values of CMC can be regarded as good.

Item-total correlation (r_{it}) refers to the correlation of an item-score to the test score, the combination of all other items within the test. The item-total correlation coefficient indicates the degree to which an item measures a similar construct as the combination of all other items (Moosbrugger, 2012b). A high value indicates that test-participants' scores on that item correlate highly to the average of the rest of the items. A low (or negative) value indicates that test-participants' scores on that item correlate marginally or even negatively to the average of the rest of the items, which indicates that the item is a weak representation of the construct. At pretest, the mean value of item-total correlation lies at r_{it} = .14 (SD = .11), with nine items with a value higher than r_{it} > .20 and four items with a negative value. At posttest, the mean value of item-total correlation lies at r_{it} = .10 (SD = .14) with seven items with a value higher than .20 and seven items with a negative value. As a rule of thumb, values of item-total correlation should be above r_{it} > .20, preferably above r_{it} > .30 (Bühner, 2011). The results of CMC for all 30 items can, therefore, be regarded as poor.

A test's internal consistency refers to the strength of the correlation between all item-scores and is calculated by an average of all possible split-half reliability-coefficients (Field, 2013). Table 10 shows the internal consistency, indicated by the Cronbach's alpha-statistic. High values indicate that all items correlate strongly with each other. Low values for internal consistency indicate that items do not measure a similar construct. At pretest, the value of item-consistency lies at α = .50.

At posttest, the value of internal consistency lies at $\alpha = .37$. Regarding internal consistency, values of .70 and higher are generally considered acceptable (Field, 2013). The results of CMC for all 30 items can, therefore, be regarded as poor (Fisseni, 2004).

7.2.3 Item-elimination.

A common strategy for the improvement of criteria for test quality lies in the elimination of items that show poor results of item-analysis (Bühner, 2011; Field, 2013) (see also chapter 6.1). As was shown in the previous section, several items showed poor values for item-total correlation. In the process of several rounds of item-elimination, the items with poor values were discarded. The process of item-elimination was performed over multiple rounds. In each round, one item was discarded from the analysis, and a new item-analysis for all the remaining items was carried out. The process of item-elimination consisted of two phases. In the first phase, items with a value lower than .10 for item-total correlation were discarded. In total, seven items were discarded. In the second phase, items were discarded when their elimination led to a net increase of internal consistency. For this phase, the values in the column labeled C*ronbach's alpha if item deleted* was used. For example, when a specific item's elimination would increase the value of Cronbach's alpha of posttest more than it would decrease the alpha of pretest, it was discarded from further analysis. In this phase, three more items were discarded.

7.2.4 Final version of CMC.

After completing the process of item-Elimination, CMC consisted of 20 items. The result of the elimination-process is shown in Table 11.

Table 11

Results of Item-Analysis of CMC-20

Item			Descriptives Pretest						Descriptives Posttest						Reliability Pretest		Reliability Posttest	
					Score Distribution						Score Distribution							
ID	Sit. Spec. Skill	CM-Dimension	*M*	*P*	0	1	2	Var(x_i)	*M*	*P*	0	1	2	Var(x_i)	r_{ir}	Alpha if Item Deleted	r_{ir}	Alpha if Item Deleted
1.1	AP	LP	0.99	.50	10	82	9	0.19	1.14	.57	7	63	20	0.28	.23	.56	.43	.52
1.3	JA	PA	1.12	.56	6	77	18	0.27	1.36	.68	2	54	34	0.28	.18	.56	.33	.53
1.6	HP	LP	0.58	.29	52	39	10	0.45	0.63	.32	42	39	9	0.44	.25	.55	.13	.56
1.8	JA	RR	1.13	.56	36	16	49	0.83	1.34	.67	19	21	50	0.66	.27	.55	.03	.57
2.2	JA	LP	0.78	.39	42	39	20	0.57	1.14	.57	22	33	35	0.62	.20	.56	.15	.56
2.3	AP	PA	1.03	.51	30	38	33	0.63	1.40	.70	13	28	49	0.54	.33	.54	.03	.57
2.4	AP	LP	1.02	.51	14	71	16	0.30	1.12	.56	10	59	21	0.33	.18	.56	.14	.55
2.5	AP	RR	0.75	.38	63	-	38	0.95	1.09	.54	41	-	49	1.00	.04	.59	.22	.54
2.6	JA	LP	1.68	.84	9	14	78	0.40	1.79	.89	5	9	76	0.28	.37	.54	.30	.54
3.1	AP	LP	1.47	.73	14	26	61	0.53	1.49	.74	8	30	52	0.43	.05	.58	.16	.55
3.3	HP	PA	1.35	.67	27	12	62	0.77	1.62	.81	14	6	70	0.55	.21	.56	.14	.56
3.5	JA	RR	1.09	.54	37	17	46	0.82	1.03	.52	39	9	42	0.91	.04	.59	.23	.54
3.6	AP	RR	1.03	.51	30	38	33	0.63	1.34	.67	8	43	39	0.41	.22	.56	.07	.56
3.7	HP	RR	1.18	.59	29	25	47	0.72	1.31	.66	22	18	50	0.71	.20	.56	.32	.52
3.8	JA	PA	1.24	.62	18	41	42	0.54	1.06	.53	26	33	31	0.64	.21	.56	.07	.57
4.1	AP	RR	0.79	.40	21	80	0	0.17	0.70	.35	33	51	6	0.35	.18	.57	.09	.56
4.2	AP	LP	1.38	.69	19	25	57	0.62	1.60	.80	10	16	64	0.47	.23	.55	.22	.54
4.3	HP	RR	1.39	.69	20	22	59	0.64	1.64	.82	8	16	66	0.41	.25	.55	.48	.51
4.4	JA	RR	0.70	.35	54	23	24	0.69	0.84	.42	46	12	32	0.85	.24	.55	.13	.56
4.6	JA	PA	1.53	.77	15	17	69	0.55	1.59	.79	14	9	67	0.56	.12	.57	.19	.55

Note. AP = Accuracy of Perception, HP = Holistic perception, JA = Justification of Action, PA = Physical Arrangement, RR = Rules and Routines, LP = Low Profile, M = Mean, P = Item-Difficulty, Var(x_i) = Item-variance, r_{it} = Item-total Correlation

At pretest, the mean values of item-difficulty is P = .56 (SD = .14), one item has a value higher than P > .75. All other 19 items' values fall within the range of .25 < P < .75. At posttest, the mean values of item-difficulty is P = .63 (SD = .16), five items have a value higher than P = .75 and all other 15 items' values fall within the range of .25 < P < .75. These results show a shift to higher values of item-difficulty as a result of the courses. Despite this shift, at pretest and at posttest, the majority of items fall within the middle range, indicating that the final version of CMC has an appropriate difficulty for measurement in the sample.

Regarding item-variance, the results of item analysis of the final version of CMC show a mean variance of $Var(x_i)$ = 0.56 (SD = 0.21) with 18 items having an item-variance above $Var(x_i)$ > .20. At posttest, mean item-variance lies at $Var(x_i)$ = 0.54 (SD = 0.21), with all 20 items having a value above $Var(x_i)$ > 0.20. Mean item-variance of the final version of CMC with 20 items has gone down the version of CMC with 30 items. This is primarily caused by the fact that five of the six single-choice items were discarded. As explained in the last section, because of the item-design, where item-scores were only 0 or 2, these items lead to relatively high values for item-variance. Regarding item-variance, the final version of CMC is considered to have good values.

The primary aim of the process of item-elimination was to improve the values for item-total correlation and, consequently, the test's internal consistency. After the elimination of 10 items, values for item-total correlation have somewhat improved: At pretest, mean item-total correlation lies at r_{it} = .20 (SD = 0.08), with 12 items having a value over r_{it} > .20 and two items having a value below r_{it} < .10. At posttest, the mean value of item-total correlation lies at r_{it} = .19 (SD = 0.12), with eight items having a value over r_i = .20 and four items having a value below r_i < .10. Notably, the two items with a low (r_{it} < .10) value of item-total correlation at pretest had a medium (r_{it} > .20) value at posttest. Additionally, the five items with a low (r_{it} < .10) value at posttest had a medium to satisfactory value at pretest (.17 < r_{it} < .33). For all these items, deleting one of these items would have led to a net reduction of internal consistency. As a result of the elimination of 10 items, internal consistency has increased at pretest from α = .45 to α = .57 and at posttest from α = .37 to α = .56.

7.3 CMC: Factor Analysis

The goal of exploratory factor analysis (EFA) lies in the identification of the factors that constitute the main construct assessed by the test. More precisely, EFA firstly shows how many factors should be identified to explain the correlations between the individual variables or items. Secondly, EFA indicates which constructs or test-dimensions these factors represent (Bühner, 2011). Therefore, EFA is often used to explore the theoretical structure of constructs within relatively new fields of research. The procedure of factor analysis consists of explaining the correlation between all items through the identification of a smaller set of factors. This means that the intercorrelation between the items is reduced when it is controlled for the influence of the identified factors (Eid et al., 2015). To explore the factorial structure of CMC, a factor analysis was performed on data of pretest (n = 101) and posttest (n = 90). EFA consists of three steps: a preliminary analysis, the choice for a method of extraction and the interpretation of the results (Bühner, 2011; Field, 2013).

7.3.1 Preliminary analysis.

The preliminary analysis explores to what degree factor analysis is expected to extract one or more factors. Regarding sample size, the literature suggests a sample size of 300 or higher, because this increases the reliability of the results. However, no exact guidelines exist and sample size can also be smaller when expected communalities are above .6 (Field, 2013). Because the sample size of CMC at pretest and posttest lies well below n = 300, the sample can be considered small for factor analysis. A second indication that factor analysis is appropriate is Bartlett's test. Bartlett's test shows whether the correlation matrix is significantly different from an identity matrix that consists of zero covariance between the included items (Field, 2013). The result of this test must be significant at the level of $p < .001$ to indicate no identity matrix exists. Analysis of the results at both pretest and posttest confirms that Bartlett's test is significant at the level of $p < .001$. A third indication is the value of the Kaiser-Meyer-Olkin's (KMO) test of sampling adequacy. Values for KMO fall between 0 and 1 and indicate to which degree measurement is appropriate for factor analysis. A value close to 0 indicates that the sum of partial correlations is large relative to the sum of correlations, which indicates that there is a low degree of shared variance and a high degree of diffusion in the pattern of the correlations. Values for the KMO-

test should be at minimum .50 to make a factor analysis appropriate, recommended are values above .60 (Field, 2013). At pretest, KMO showed a value of .473, and for posttest a value of .516. These results indicate that for CMC, sampling adequacy is poor and a factor analysis might be inappropriate.

In the correlation matrix, Pearson's *r* correlation coefficients between each of the variables are presented. Table 12 shows the correlation matrix of all 20 items included in the final version of CMC. Values of pretest are placed under the diagonal; values of posttest are placed above the diagonal. These results show an uneven distribution of item-correlation with mostly low and non-significant values. At pretest, only 3 of 180 correlation coefficients were significant at the .05 level. At posttest, 35 of 180 correlation coefficients were significant at the .05 level. The low values within the correlation matrix indicate that the items do not share much variance. These results correspond to the low values of item-total correlation and low values of internal consistency reported in chapter 7.2.2 and the low value of the KMO test. These combined values represent an indication that a factor analysis is not appropriate to identify underlying factors within the data. Because the combined result of the KMO test of pretest and posttest lie around the minimum value of .500, factor analysis is nevertheless calculated. Interpretation of results, however, should take these values from the preliminary analysis into account.

Table 12

Correlation Matrix for 20 Items of CMC on Pretest (below the diagonal) and Posttest (above the diagonal)

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1. 1.1	–	.256*	.216*	.144	.111	.023	.162	.039	.309*	.085	.140	.079	-.049	.250*	.060	.140	.254*	.350*	.184*	.066
2. 1.3	-.091	–	.379*	.079	.092	-.023	.040	-.061	.192*	.077	.031	.111	.100	.229*	-.021	.022	.275*	.212*	.115	.090
3. 1.6	-.014	.189*	–	-.055	.060	-.088	-.058	-.069	.097	-.203*	.172	-.087	.090	.146	-.046	.061	.070	.060	.145	.123
4. 1.8	.054	.149	.023	–	.027	.030	-.115	.003	.093	-.130	-.024	.043	.050	.055	.022	-.111	.150	.044	-.123	.051
5. 2.2	.115	.073	-.122	.070	–	.035	-.138	.112	-.088	.014	-.021	.113	.079	.118	-.031	-.099	.025	.192*	.170	.064
6. 2.3	.030	.123	.061	.160	.161	–	.016	.012	.133	.150*	-.091	.093	-.154	-.022	-.038	-.031	-.013	.115	-.073	.119
7. 2.4	.127	.145	.269*	.055	.228*	.045	–	.078	.012	.107	.109	-.007	-.176	.221*	.156	.010	.211*	.210*	.057	.040
8. 2.5	.018	-.109	.055	.002	-.047	.100	.009	–	.057	.104	.076	.079	.215*	-.060	.148	.179*	.249*	.085	.137	.004
9. 2.6	-.121	.260*	.183*	.228*	.126	.199*	-.097	-.031	–	.171	-.005	.103	.118	.098	-.105	.191*	.012	.471*	.047	.204*
10. 3.1	.141	-.075	.155	-.106	.004	.028	.052	.037	-.046	–	-.193*	.153	.156	.270*	.076	.035	.040	.150	-.132	.116
11. 3.3	.140	.020	.112	.006	-.081	.043	.027	.183*	.182*	-.036	–	.240*	.064	.100	.187*	-.108	-.168	.069	.159	.041
12: 3.5	.205*	-.094	-.037	.143	-.030	.163	-.084	-.190*	-.072	.027	.087	–	.018	.211*	-.002	.158	-.117	.111	-.020	.193*
13. 3.6	-.115	.176*	-.014	.064	.161	.158	.137	.178*	.099	.166*	.028	-.059	–	.007	-.016	.039	.113	.083	-.137	-.124
14. 3.7	.112	.120	.131	.189*	.061	.051	.099	.126	.162	-.054	.131	.044	.007	–	-.009	.009	.101	.290*	-.009	.152*
15. 3.8	.070	.061	.020	.162	.148	-.012	.038	.194*	.292*	.164	.057	-.092	.142	.059	–	-.012	.124	.039	.088	-.225*
16. 4.1	.101	.026	.194*	.126	-.051	.019	.019	-.005	.052	.026	.008	.024	.081	.079	-.034	–	.061	.042	.037	-.028
17. 4.2	.128	.147	.053	.113	.173*	.287*	-.064	.018	.102	.023	.070	.051	.142	-.071	-.035	.153	–	.210*	.025	-.062
18. 4.3	.269*	.036	.060	.123	.140	.202*	.234*	-.120	.245*	-.037	.050	.035	.092	-.029	.148	.249*	.069	–	.114	.277*
19. 4.4	.213*	-.036	.172*	.077	.166*	.074	.035	.032	.238*	-.215*	.252*	.035	.044	.118	-.047	.022	.035	.144	–	.102
20. 4.6	.017	.102	.272*	-.132	.031	.126	-.002	-.230*	.258*	.145	.035	.107	-.061	.022	-.033	.074	.149	-.065	.146	–

Note. Item 1 to 20 are marked with their item-ID

* $p < .05$

7.3.2 Method of extraction.

In the next step, a method for the extraction of factors and rotation is chosen. There are several methods to extract factors through EFA. Because the goal of the factor-analysis was to explore the underlying factor structure, and not the reduction of data, the method of Principal Axis Factoring (PAF) was preferred over the principal component analysis (PCA) (Bühner, 2011). Additionally, a factor rotation was chosen to check whether better factor loadings could be calculated by rotating the axes. Because it was expected that factors correlated, the oblique Promax rotation method was chosen (Field, 2013).

7.3.3 Interpreting the data.

The third step consists of interpreting the results of the factor analysis. The following concepts are relevant for the understanding of the results of the method of principal axis factoring (Bühner, 2011). The *factor loading* refers to the relative contribution that a variable makes to a factor. The value of a factor loading represents the change in the number of standard deviations on the item-score when the factor value is changed by one standard deviation. A high value means an item 'loads' strongly on that factor, indicating that this item, together with other items with a similarly high value, represents one shared factor. In EFA, values for *commonalities* and *eigenvalues* are often reported. However, these parameters are only relevant for orthogonal rotation techniques that are used for uncorrelated factors. Because the current study uses the oblique rotation technique for correlated factors, commonalities and eigenvalues were not reported. In oblique rotation, a *pattern matrix* and a *correlation matrix were* calculated. In a pattern matrix, the unique contribution of an item to a factor is portrayed. The *correlation matrix* portrays the correlation coefficients between the factors as predicted by the structure matrix.

Table 13

Summary of Factor Loadings for Promax Nine Factor Solution for CMC Pretest

Item			Factor Loadings								
ID	Sit. Spec. Skill	CM-Dimension	1	2	3	4	5	6	7	8	9
2.6	JA	LP	**.93**	.10	-.06	.00	.18	-.02	.00	.00	.10
3.8	JA	AC	**.39**	.05	.17	-.09	-.04	-.18	.12	.33	.15
4.3	HP	RR	.13	**1.01**	-.05	.01	.02	.05	-.04	-.04	-.10
4.1	AP	RR	-.05	**.26**	.00	.19	-.02	.17	-.22	.02	.09
2.5	AP	RR	-.06	-.09	**.91**	.02	.29	.13	-.09	.04	-.03
4.6	JA	AC	.21	-.14	**-.32**	.23	.17	.14	.04	.18	-.21
3.5	JA	RR	-.13	-.05	**-.28**	-.10	.10	.14	-.09	.08	.25
1.6	HP	LP	.05	.03	.03	**.74**	.16	-.02	-.22	.10	.01
2.4	AP	LP	-.25	.10	.01	**.50**	.00	-.13	.37	.01	.06
1.3	JA	AC	.17	-.06	-.08	**.27**	-.18	.15	.10	-.09	.20
4.4	JA	RR	.13	.02	.05	.06	**.68**	.00	.18	-.29	-.08
3.3	HP	AC	.11	.00	.19	.08	**.42**	.01	-.09	-.05	.02
1.1	AP	LP	-.22	.15	-.07	-.06	**.40**	.00	.12	.21	.11
4.2	AP	LP	-.07	.03	.06	-.04	.00	**.64**	.09	.01	-.01
2.3	AP	AC	.03	.07	.08	.00	.03	**.48**	.10	-.01	.08
3.6	AP	RR	.05	.04	.27	.04	-.15	**.27**	.20	.10	-.01
2.2	JA	LP	.05	-.08	-.06	-.17	.09	.16	**.75**	.01	-.04
3.1	AP	LP	.00	-.04	.01	.10	-.24	.03	.00	**.79**	-.11
1.8	JA	RR	.12	-.03	-.07	.01	-.08	.10	-.05	-.09	**.68**
3.7	HP	RR	.08	-.17	.04	.18	.23	-.11	.02	-.04	**.35**
			Factor Correlation								
Factor 1			–								
Factor 2			-.05	–							
Factor 3			.05	.02	–						
Factor 4			.24	.04	.00	–					
Factor 5			-.09	.16	-.23	-.04	–				
Factor 6			.25	.00	-.23	.13	.18	–			
Factor 7			.07	.34	.16	.19	-.09	-.04	–		
Factor 8			-.13	.14	-.02	-.02	.31	.10	.01	–	
Factor 9			.00	.36	.16	-.04	.18	.05	.24	.12	–

Note. Item 1 to 20 are marked with their item-number and the acronyms of the theoretical dimensions: AP = Accuracy of Perception; HP = Holistic Perception; JA = Justification of Action; PA = Physical Arrangement of the Classroom; RR = Rules and Routines, LP = Low Profile.

Table 14

Summary of Factor Loadings for Promax Eight-Factor Solution for CMC-Posttest

Item			Factor Loadings							
ID	Sit.- Spec. Skill	CM-Dimension	1	2	3	4	5	6	7	8
2.6	JA	LP	**.97**	.04	-.01	-.04	-.18	.08	-.23	.05
4.3	HP	RR	**.40**	.07	.02	.21	.10	-.06	.14	.06
4.1	AP	RR	**.30**	.03	-.07	-.04	.07	.09	-.07	-.20
1.1	AP	LP	**.25**	.24	.06	.04	.20	-.10	.04	.10
1.6	HP	LP	.04	**.78**	.05	-.13	-.19	.11	.01	-.16
1.3	JA	CA	.05	**.55**	-.05	.14	-.01	.11	.04	.03
2.3	AP	CA	.15	**-.20**	-.09	.10	-.07	-.17	.06	.07
3.3	HP	CA	-.04	.04	**.98**	-.00	.17	.09	-.07	.05
3.1	AP	LP	.06	-.13	-.14	**.66**	.06	.18	-.02	-.20
3.7	HP	RR	-.13	.29	.10	**.57**	.04	-.01	.03	.01
3.5	JA	RR	.07	-.12	.26	**.32**	-.09	.06	.14	.03
4.2	AP	LP	-.03	.19	-.26	-.05	**.58**	.08	.01	.18
2.4	AP	LP	-.02	-.03	.09	.22	**.49**	-.28	-.22	-.13
3.8	JA	CA	-.12	-.13	.20	-.01	**.44**	.02	-.05	.01
2.5	AP	RR	.15	-.25	.06	-.07	**.39**	.24	.23	-.08
4.6	JA	CA	.21	.05	.01	.17	**-.25**	-.18	.17	.00
3.6	AP	RR	.07	.17	.10	.14	-.00	**.81**	.01	.01
2.2	JA	LP	-.23	.04	-.07	.11	-.13	.05	**.73**	.05
4.4	JA	RR	.09	.11	.07	-.23	.12	-.20	**.32**	-.21
1.8	JA	RR	.02	-.10	.06	-.15	.02	.02	.01	**.74**
			Factor Correlation							
Factor 1			–							
Factor 2			.26	–						
Factor 3			.05	.12	–					
Factor 4			.35	.11	-.04	–				
Factor 5			.26	.27	-.08	.13	–			
Factor 6			-.11	-.21	-.17	-.18	-.01	–		
Factor 7			.36	.22	.15	.11	.24	-.04	–	
Factor 8			.17	.24	-.12	.24	.08	-.03	.08	–

Note. Item 1 to 20 are marked with their item-number and the acronyms of the theoretical dimensions: AP = Accuracy of Perception; HP = Holistic Perception; JA = Justification of Action; PA = Physical Arrangement of the Classroom; RR = Rules and Routines, LP = Low Profile.

As described in chapter 6.2.1, the items related to the construct of classroom management competence were formulated based on the combination of two theoretical dimensions: the dimension of classroom management: physical arrangement of the classroom (AC), rules and routines (RR), and Low Profile (LP), and the dimension of classroom management expertise: accuracy of perception (AP), holistic perception (HP), and justification of action (JA). Items were formulated in a way that at least two items represented each of the nine possible combinations (see chapter 6.2.3 and Table 5). Item-elimination based on the results of item-analysis (see chapter 7.2.3) eliminated a total of 10 items. After elimination, each of the three aspects of both dimensions was represented by at least four items. Three of the nine combinations of aspects were only represented by one item (see Table 15). Because of this uneven distribution of items over the nine theoretical factors, it was expected if factor analysis would be able to identify sub-scales, it would only be able to show one of the two dimensions, either the three dimensions related to classroom management (PA, RR, LP), or the three dimensions related to classroom management expertise (AP, HP, JA).

Table 15
Distribution of Items across the Theoretical Dimensions

Theoretical Dimensions	Classroom Arrangement (CA)	Rules and Routines (RR)	Low Profile (LP)	Total
Accuracy of Perception (AP)	2.3	2.5; 3.6; 4.1	1.1; 2.4; 3.1; 4.2	8
Holistic Perception (HP)	3.3	3.7; 4.3	1.6	4
Justification of Action (JA)	1.3; 3.8; 4.6	1.8; 3.5; 4.4	2.2; 2.6	8
Total	5	8	7	20

The result of EFA is presented in Table 13 for pretest and Table 14 for posttest. At pretest, nine factors were extracted and at posttest, eight factors were extracted. Items with factor loadings of .45 or higher were distributed across all factors. The items that were grouped according to the value of the factor loadings could not be allocated to any of the theoretical dimensions (see Table 15). Also, the items were grouped differently between pretest and posttest. These results of EFA

confirm the results from item-analysis (see chapter 7.2.2) and internal consistency (see chapter 7.2.4) and indicate that no valid factors can be extracted from the items of CMC.

7.4 Descriptive Statistics

In the following sections, descriptive results of all individual measurements and variables are presented. For each of the variables, the measurement, its constituent variables and the function of the values are summarized. Additionally, for each of the measurements, the number of missing values, as well as the method with which they were accounted for in further analysis, are reported. Mean values and standard deviations are reported to summarize the result of descriptive statistics for each variable. Results are provided for each group and measurement time. Groups are indicated with a number (1 or 2) indicating the cohort and an abbreviation (Sit. or Cog.) indicating the instructional strategy of the course.

7.4.1 CMC.

The video test CMC, measuring classroom management competence was developed to answer the main research question and to evaluate the effectiveness of the video-based courses. The video test was administered at pretest two months before the course and at posttest on the last day of the course.

Examination of the values of CMC-score showed at pretest one outlier with a value lower than the 25th percentile -1.5 x IQR (case 91). At posttest, two outliers (case 16 and case 91) had a value lower than the 25th percentile -1.5 x IQR and one case had a value higher than the 75th percentile + 1.5 x IQR (case 5). Outliers can severely bias procedures of data-analysis based on the estimation of parameters and their spread (Field, 2013). An outlier can be trimmed from the data set when it is expected to be the result of a case not belonging to the population, being the result of a mistake or the test participant's failed interpretation or boycott of the item (Eid et al., 2015). Trimming is typically done using a percentage-based rule or a standard deviation-based rule. The percentage-based rule refers to trimming the upper and lower cases (for example 5%) of the distribution. The standard deviation based rule refers to trimming those cases that are 2.5 standard deviations above and below the sample mean (Field, 2013). Inspection of the outliers

of CMC showed that case 91 had a value of 5 at pretest and a value of 9 at posttest, both well below 2.5 standard deviations of the respective means. Moreover, case 91 also has outlying values at pretest and posttest on CME (see chapter 7.4.2) and received a missing value of FAIR-2 as a result of the scoring guidelines of FAIR-2 that assume a failed understanding of the test-instructions (see chapter 7.4.3). It was decided to trim case 91 from data-analysis related to construct validity and the evaluation of the effectiveness of the courses. It was expected that this case received a low score due to a failed understanding of the test-instructions. The values for the other outliers of CMC (case 5 and case 16), fell within the range of 2.5 standard deviations below or above the mean. Therefore, they were not trimmed and included in the sample for further analysis. Table 16 gives an overview of the mean test-score and standard deviation of CMC for each group.

Table 16
Mean Test-Scores and Standard Deviations for CMC

	1Sit.[a]		1Cog.[b]		2Sit.[c]		2Cog,[d]	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Pretest	22.1	5.0	24.2	5.3	22.0	5.1	21.7	3.9
Posttest	24.8	5.0	26.0	5.0	25.6	3.4	25.5	4.7

Note. 1Sit. = group from the first cohort in the course with situated strategy; 1Cog. = group from the first cohort in the course with cognitive strategy; 2Sit. = group from the second cohort in the course with situated strategy; 2Cog. = group from the second cohort in the course with cognitive strategy.
[a]*n* = 24. [b]*n* = 20. [c]*n* = 20. [d]*n* = 23.

7.4.2 CME.

The video test CME (König, 2015b) was administered in parallel to the new video test CMC to measure its convergent validity (see chapter 6.4). Like CMC, CME's item-construction is based on the situation-specific skill of accuracy of perception (AP), holistic perception (HP) and justification of action (JA). In a validation study, confirmatory factor analysis showed that CME should be considered a one-dimensional construct (König, 2015b). Therefore, the single test-score consisting of the added scores of all 24 items is reported. Calculation of internal consistency on all 24 items showed a value of α = .45 at pretest and α = .52 at posttest, which can be considered poor values (Fisseni, 2004). Moreover, they are distinctly lower than the value of α = .76 reported

in the validation studies of CME (König, 2015b).

As described in the previous section, missing values were entered when participants did not reach all questions in the available time. Exploration of the missing values showed that at pretest 11 (0.45%) and at posttest 6 (0.28%) of total values were missing. At both pretest and posttest, these missing values occurred in four of the 24 items of both pretest and posttest. Missing values occurred in seven of 101 (6.9%) participants at pretest and four of 90 (4.5%) participants at posttest. To be able to include these cases for further analysis, equally to CMC, it was decided to use the method of multiple imputation to process missing values. After the calculation of multiple imputation, the median value for all five imputations was calculated and fed back into the data-matrix to create one single value for each of the missing values.

As reported in the preceding section, the examination of the values of CME showed the case 91, as an outlier in both pretest and posttest with a value below the 25th percentile -1.5*IQR. As case 91 was trimmed following the systematic discussed in the preceding section, the distribution showed no other outliers for CME. Table 17 shows the mean scores and its standard deviations for CME for each of the groups and each of the cohorts.

Table 17
Mean Test-Scores and Standard Deviations for CME

Variable	1Sit.[a]		1Cog.[b]		2Sit.[c]		2Cog.[d]	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Pretest	15.1	2.9	15.2	2.9	14.5	1.6	15.5	2.8
Posttest	16.1	2.7	15.1	2.3	14.4	2.9	15.1	2.5

Note. 1Sit. = group from the first cohort in the course with situated strategy; 1Cog. = group from the first cohort in the course with cognitive strategy; 2Sit. = group from the second cohort in the course with situated strategy; 2Cog. = group from the second cohort in the course with cognitive strategy.
[a]*n* = 24. [b]*n* = 20. [c]*n* = 20. [d]*n* = 23.

7.4.3 FAIR-2.

The instrument FAIR-2 (Moosbrugger & Oehlschlägel, 2011) was used to measure participating prospective teachers' attentiveness (see also chapter 3.5.5 and chapter 6.4.2). FAIR-2 was

administered in the morning on the second day of the course with the purpose to assess discriminant validity of CMC. FAIR-2 creates three scores: performance (P), quality (Q), and continuity (C) (see chapter 6.4.2). One participant arrived too late in the course and, therefore, could not participate in the FAIR-2 test. The coding software (Moosbrugger et al., 2017) returned missing values for three additional participants. Although these participants filled out the questionnaire successfully, they marked more than 5% of the processed items invalidly. The FAIR-2 scoring guidelines indicate that for such cases, a missing value should be submitted because this indicates that a test-participant did not understand the test-instructions. For the final sample of 88 participations, a total of four participants had missing values. Because these missing values represent less than 5% of the cases (Field, 2013), data analysis was performed using list-wise exclusion.

Examination of the values of FAIR-2 showed two outliers for the P-score (<1.5*IQ) and the two outliers for the C-score (>1.5*IQ and <1.5*IQ). Because cases for which it is expected that they did not understand test-instruction were already excluded, no further measures for bias-reduction were undertaken. The inspection of the values of Q-scores showed strong skewness. These skewed distributions are the result of the calculation method and therefore do not represent unusual values (Moosbrugger & Oehlschlägel, 2011). Table 18 shows the mean scores and its standard deviations for all three scores on FAIR-2 for each of the groups.

Table 18
Mean Test-Scores and Standard Deviations for FAIR-2

	1Sit.[a]		1Cog.[b]		2Sit.[c]		2Cog.[d]	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
FAIR-2, P-score	369.9	91.4	334.2	74.7	372.6	58.1	389.4	102.68
FAIR-2, Q-score	0.952	0.059	0.965	0.030	0.958	0.031	0.965	0.026
FAIR-2, C-score	345.5	95.4	322.3	72.2	356.9	56.6	375.6	99.2

Note. 1Sit. = group from the first cohort in the course with situated strategy; 1Cog. = group from the first cohort in the course with cognitive strategy; 2Sit. = group from the second cohort in the course with situated strategy; 2Cog. = group from the second cohort in the course with cognitive strategy.
[a]*n* = 23. [b]*n* = 18. [c]*n* = 20. [d]*n* = 23.

7.4.4 Scales of evaluation project of the practical semester.

All prospective teachers in the study participated in the practical semester, either before or after their course on classroom management. As such, they were all questioned as part of the full survey held in the related evaluation project (Practical Semester Hesse, 2018). From the data assessed in the evaluation study of the practical semester, the following data were used (see chapter 6.4.3): a scale concerning self-evaluation of classroom management skills (Pfitzner-Eden et al., 2014) consisting of four rating-items on a Likert-scale related to dealing with classroom disruptions. The values of these items were added to form a scale on self-efficacy on classroom management ranging between 4 and 24. Also, the values of two additional items that related to general cognitive ability were used: the self-reported grade from the introductory university-course of Bildungswissenschaften (BW-A), ranging between 1 and 15, and the self-reported grade of the high-school diploma, ranging between 4.0 and 1.0 (1.0 being the highest possible score and 4.0 the lowest possible score). Both scales were used for the assessment of convergent validity.

The data of 78 of the 88 course-participants of the final sample were successfully matched to the data of the practical semester. The pseudonyms of the other ten course-participants could not be found within the data of the practical semester. Four participants of which the data could be matched showed missing values for all the above-listed variables. Of the remaining 74 prospective teachers, valid data could be successfully matched. Additionally, some test-participants left out specific survey-items, leading to additional missing values. Because the information in these items was not assumed to be related to issues of privacy (Döring & Bortz, 2016b), no bias in the answering of these items was assumed, and missing values were deleted pairwise in further data analysis.

Exploration of the values showed that one participant entered a value of 6.0 for the grade of his high-school diploma. Because this grade would imply, this particular participant would not have graduated from high school, a missing value was imputed instead. Further exploration revealed two other outliers with values of 3.0. However, because these values are valid for this variable, no further changes were made. No further outliers were found on the other variables.

Calculation of internal consistency for the scales of self-efficacy on classroom management showed a value of α = .91, which means these scales can be considered to be of high reliability. This value is equivalent to values reported in validation studies (Pfitzner-Eden et al., 2014). Table 19 presents the results of all the scores for each of the groups.

Table 19
Mean Values and Standard Deviations of Self-Efficacy, High-School, and BW-A grades

	1Sit.			1Cog.			2Sit.			2Cog.		
Variable	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Self-Efficacy	15	17.0	2.8	14	17.9	4.3	14	12.9	2.6	17	13.4	5.0
High-school grade	20	1.9	0.4	16	2.0	0.4	14	1.9	0.5	15	1.9	0.3
BW-A Grade	16	10.2	2.4	17	10.1	2.0	11	11.6	1.6	18	11.8	3.0

Note. 1Sit. = group from the first cohort in the course with situated strategy; 1Cog. = group from the first cohort in the course with cognitive strategy; 2Sit. = group from the second cohort in the course with situated strategy; 2Cog. = group from the second cohort in the course with cognitive strategy.

7.4.5 Course-evaluation.

To investigate the subjective experience of the participating prospective, standard items of the course-evaluation questionnaire of Goethe University Frankfurt were used (see chapter 6.4.5; LUQ, 2018). Results of these rating-items were used to explore to what degree differences between the two cohorts existed. Additionally, it was explored to what degree the situated and cognitive strategies led to differences in the subjective experiences with the learners, for example as a result of differences in motivation. Rating-items consisted of Likert-scales regarding gains in knowledge, course-structure, course-pace, and the quality of learning materials (see chapter 6.4.5).

Of the 88 prospective teachers of the final sample, the data of one participant could not be matched to the main data-set. Additionally, some missing values occurred because participants did not fill out a specific item. In Table 20 for each of the rating-items, the number of valid answers is indicated. Scores on the 6-point Likert-scales ranged from 1 does not apply to 6 applies. Hence, scores can be considered of interval level (Field, 2013). However, scores were skewed. Table 20 summarizes the results of the descriptive statistics of all participants.

Table 20
Mean Values and Standard Deviations of Course-Evaluation

Item	1Sit.[a]		1Cog.[b]		2Sit.[c]		2Cog.[d]	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
1. Course-participation leads to a noticeable increase in knowledge.	3.7	1.1	4.1	1.3	4.9	1.2	5.4	0.8
2. The learning pace of the course is appropriate.	4.7	1.8	4.3	1.4	5.3	1.2	5.1	1.3
3. The subject matter of the course is well structured.	4.6	1.1	5.2	0.8	5.3	1.0	5.3	1.2
4. Materials and media are used appropriately in the course.	5.5	0.9	5.7	0.6	5.7	0.6	5.6	1.1

Note. 1Sit. = group from the first cohort in the course with situated strategy; 1Cog. = group from the first cohort in the course with cognitive strategy; 2Sit. = group from the second cohort in the course with situated strategy; 2Cog. = group from the second cohort in the course with cognitive strategy.
[a]n = 24. [b]n = 20. [c]n = 20. [d]n = 22.

7.5 CMC: Reliability and Construct Validity

As an indicator of the reliability of CMC, internal consistency and the correlation coefficient between pretest and posttest were calculated. As was reported in chapter 7.2.4, CMC showed a value for internal consistency of α = .57 at pretest and α = .56 at posttest. As a general guideline, values for internal consistency should fall between .70 < α < .90 (Fisseni, 2004). The values for CMC fall below these general rules of thumb, indicating poor reliability. Additionally, reliability for CMC is calculated by correlating the scores of pretest and posttest. There were around two months between pretest and posttest. This period was chosen as it was assumed to be long enough to reduce the learning effect of participation from pretest, but also short enough to reduce the probability of confounding influences on the posttest. Between pretest and posttest, prospective teachers participated in the video-based courses, meaning that the measurements were not held under similar conditions. As such, the value of the correlation coefficient does not represent test-retest reliability. However, still, a value of at least moderate magnitude between pretest and positive should be expected for a reliable instrument. Because of the sample size (n = 87), it is assumed that the central limit theorem guarantees a normally distributed sampling distribution and Pearson's r was used as a correlation coefficient (Field, 2013). CMC's scores at pretest and posttest were significantly related, r = .37, p < .001. The value of Pearson's r can be considered of medium magnitude (Cohen, 1988). However, a value of r = .37 as an indicator of

reliability should be considered poor.

In chapter 6.7, the following hypotheses for construct validity were formulated. Regarding convergent validity, it was expected that scores for CMC correlate positively with a similar video of CME at both pretest and posttest (H1a). Regarding discriminant validity, it was expected that for the second cohort, posttest-scores of CMC do not correlate with scales for self-efficacy at the beginning of the practical semester (H1b). Also, it was expected that scores of CMC do not correlate with scores of FAIR-2 (H1c) or with scales of general cognitive ability, like the grade of the BW-A-module or the grade of the high-school diploma (H1d). It is assumed that the sample size (n = 87) ensures that the central limit theorem is applicable and, therefore, that the assumption of a normally distributed sampling distribution is met and Pearson's r can be used as a correlation coefficient (Field, 2013).

Analysis showed that scores for CMC were significantly related to scores on CME at pretest, r = .40, p < .001 and at posttest, r = .22, p = .04, confirming H1a regarding convergent validity. Scores for performance (P) of FAIR-2 were neither related to scores at pretest of CMC, r = .10, p = .37 nor to scores at posttest of CMC r = .12, p = .28. Scores for continuity (C) of FAIR-2 were neither related to scores at pretest of CMC r = .14, p = .21 nor to scores at posttest r = .13, p = .26. However, analysis showed that scores for quality (Q) of FAIR were related to CMC's scores at pretest, r = .31, p = .004 of CMC, but not to scores at posttest, r = .17, p = .12. These results mostly confirm H1b regarding discriminant validity. Scores for posttest on CMC of the second cohort were not related to scores on scales for classroom management self-efficacy r = -.21, p = .25, confirming H1c regarding discriminant validity. Grades on BW-A were neither related to scores at pretest of CMC r = -.04, p = .77 nor to scores at posttest of CMC r = -.03, p = .79. Grades on the high-school diploma were neither related to scores at pretest of CMC r = .07, p = .59 nor to scores at posttest of CMC r = -.03, p = .80, confirming hypothesis H1d regarding discriminant validity.

7.6 Course-Evaluations

Mean values (see Table 20) on the four items of the course-evaluations were compared for three purposes (see chapter 6.4). Firstly, to investigate to what degree prospective teachers with

practical experiences found the course easier than the course-participants without practical experience. Secondly, to investigate the degree to which prospective teachers in the course following a cognitive strategy perceived their course to be better structured than the prospective teachers in the course following a situated approach. And, thirdly, to investigate to what degree differences in prospective teachers' perceived use of materials and media differed between the two course-types. Mean comparison was performed twice for all four items, once to compare means between cohort one and cohort two and once to compare means between the cognitive and the situated instructional approach. One-way ANOVA with Bonferroni correction for the family-wise error was used (Field, 2013). With a total of eight separate mean comparisons, the corrected critical p-value was $p = .006$ (Eid et al., 2015).

Regarding the comparison of mean differences between the two cohorts with the Bonferroni correction, the results show that course-participants from the second cohort without experience in the practical semester perceived their course as leading to higher degrees of knowledge-gains (item 1) with a large effect size, $F(1, 84) = 29.22$, $p < .001$, $d = 1.18$. Without the Bonferroni correction, participants from the second cohort perceived their course having a more appropriate pace (item 2), with a medium effect-size $F(1, 84)^{8} = 7.23$, $p = .009$, $d = 0.58$. Furthermore, they perceived their course as having a higher degree of structure (item 3), with a medium effect-size $F(1, 84) = 6.89$, $p = .010$, $d = 0.58$. Mean values of perceived use of materials and media (item 4) did not significantly differ between the two cohorts $F(1, 84) = 1.77$[8], $p = .188$.

These results indicate that prospective teachers from the second cohort perceived the course to contribute considerably more to knowledge gains than prospective teachers from the first cohort. This could indicate that the course was too easy for the first cohort. This explanation is confirmed by the fact that prospective teachers from the second cohort rated the pace of the course as more appropriate than the prospective teachers from the first cohort. As this last result was only significant without the Bonferroni correction of the critical p-value, there is an increased probability that this mean-difference occurred by chance alone. An alternative explanation of

[8] Because of significant value for Levenes' test for Homogeneity of Variances, results of Welch's F-Test for robust Equality of means are reported.

this difference between the two cohorts could be that practice effects occurred and the courses of the second cohort were qualitatively better than from the first cohort. Indeed, the courses were not piloted, so the possibility of practice effects cannot be excluded. However, as learning materials and the sequence and instruction of exercises were standardized, practice-effects are expected to be limited, indicating that the courses of the prospective teachers of the first cohort with experience in the practical semester were relatively easy.

Regarding the comparison of mean differences between the two course-types, no significant mean differences were found with the Bonferroni correction of the critical p-value. Without the Bonferroni correction, participants in the courses with a cognitive strategy perceived their course as leading to higher degrees of knowledge-gains (item 1) with a medium effect-size, $F(1, 84)$[9] = 4.06, $p = .047$, $d = .044$ and having a higher degrees of structure (item 3), with a medium effect-size, $F(1, 84)$[9] = 5.63, $p = .020$, $d = .62$. No differences were found regarding the perceived appropriateness applied materials and media (item 4), $F(1, 84) = 1.71$, $p = .194$ and having a more appropriate pace with a medium effect-size (item 2), $F(1, 84) = 0.24$, $p = .627$, $d = 0.58$.

These results indicate that perceptions of prospective teachers regarding the courses was largely similar. However, the participants in the course following the cognitive strategy rated knowledge-gains, degree of structure and perceived pace significantly higher, when the Bonferroni-correction was not applied. Taking into account the increased probability of a type I-error when interpreting p-values without a correction for family-wise error, these results indicate that participants slightly preferred the course with the cognitive instructional design. Because also without the Bonferroni-correction, the mean difference regarding the use of learning materials and media indicate that in both courses, video clips were perceived to be applied appropriately.

[9] Because of significant value for Levenes' test for Homogeneity of Variances, results of Welch's F-Test for robust Equality of means are reported.

7.7 Treatment Effects

7.7.1 Effects of video-based courses on self-efficacy.

In chapter 6.7, the hypothesis was formulated that because prospective teachers of the second cohort participated in a video-based course before starting the practical semester, their scores on perceived self-efficacy on classroom management would be higher than scores of the first cohort (H2a). Levene's test for equality of error variance showed a non-significant result ($F(1, 58) = .54$, $p = .46$), indicating that the assumption was met. Independent-samples t-test showed that on average, prospective teachers from the second cohort had lower scores on perceived classroom management self-efficacy ($M = 13.2$, $SE = 0.73$) than prospective teachers from the first cohort ($M = 17.4$, $SE = 0.66$). This difference was statistically significant $t(58) = 4.31$, $p < .001$ and had a large effect-size $d = 1.12$. This result shows the opposite effect as was hypothesized.

7.7.2 Effects of video-based courses on situation-specific skills.

Before testing the hypothesis regarding the main effect of the video-based courses, it is firstly tested, if there is a meaningful difference in classroom management competence between the two cohorts at pretest. For this, an independent t-test was carried out which compared the mean values of the two cohorts at pretest. Results show that, on average, prospective teachers with experience in the practical semester have higher scores on CMC ($M = 23.1$, $SE = 0.79$) than prospective teachers without experience in the practical semester ($M = 21.8$, $SE = 0.68$). This difference was not significant $t(85) = 1.16$, $p = .25$, indicating that there is insufficient evidence the mean difference can be explained by the participation in the practical semester.

To test the hypotheses regarding the effectiveness of the video-based courses on situation-specific skills (H2b-H2f), a mixed-design ANOVA was calculated consisting of a two-way ANOVA with repeated-measurement ANOVA. For the calculation of a mixed-design ANOVA, two conditions should be met: firstly, residuals for each of the groups in measurements of both pretest and posttest should be normally distributed. Secondly, error variance should be equal across the groups. The third assumption of sphericity is only valid within models with three or more repeated measurements (Field, 2013) and is therefore not tested. Levene's test for equality

of error variances showed non-significant results for both pretest ($F(3, 83) = 0.83$, $p = .479$) and posttest ($F(3, 83) = 1.08$, $p = .363$), indicating that the first assumption was met.

As the sample sizes within the groups were too small for the central limit theorem to take effect, the assumption of normally distributed residuals cannot be taken for granted and must be tested. For this, Q-Q plots (see Appendix D), as well as Shapiro-Wilk tests, were calculated for each group on both pretest and posttest. Results showed non-significant values for all measurements on all groups, except for the posttest of group 1Cog. ($p = .049$). Deviations of normally distributed residuals tend to run the risk of invalid p-values. This leads to the increased probability of type I error, falsely rejecting a true null-hypothesis, and type II error, falsely retaining the null-hypothesis. Because no non-parametric tests are available in SPSS 23 and the deviations of normality only occurred for one group at posttest, it was decided to run the ANOVA. However, Interpretation of the results of the ANOVA should take these limitations into account, especially when p-values of the found effects are close to the threshold of .05.

Results of the mixed-design ANOVA showed that, firstly, there was a significant main effect of course-participation on CMC-scores with a large effect size, $F(1, 83) = 27.14$, $p < .001$, $\eta^2 = .246$, confirming hypothesis H2b (see Figure 10.).

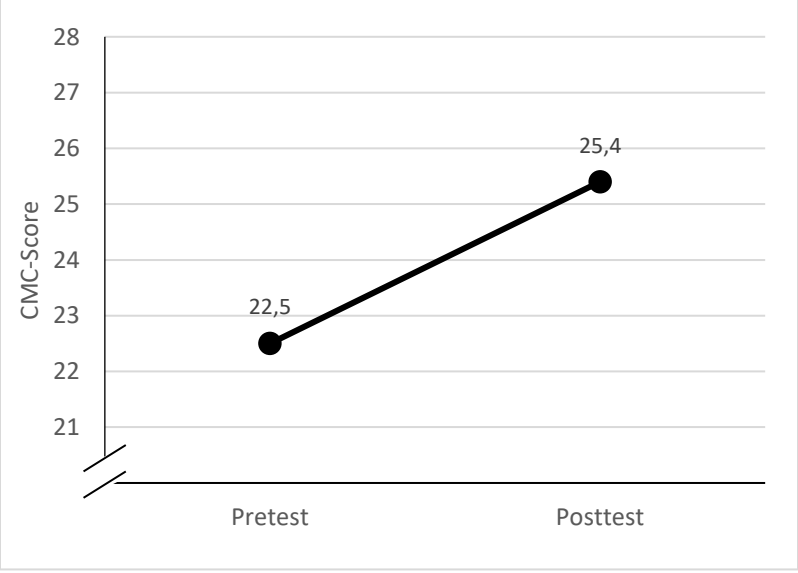


***Figure 10*. Mean Change between Pretest and Posttest for the whole sample**

Secondly, there was no significant first-order interaction effect of cohort, indicating that change in CMC-scores did not depend on practical experience, ($F(1, 83) = 1.69$, $p = .20$, $\eta^2 = .020$) and confirming hypothesis H2c (see figure 11.).

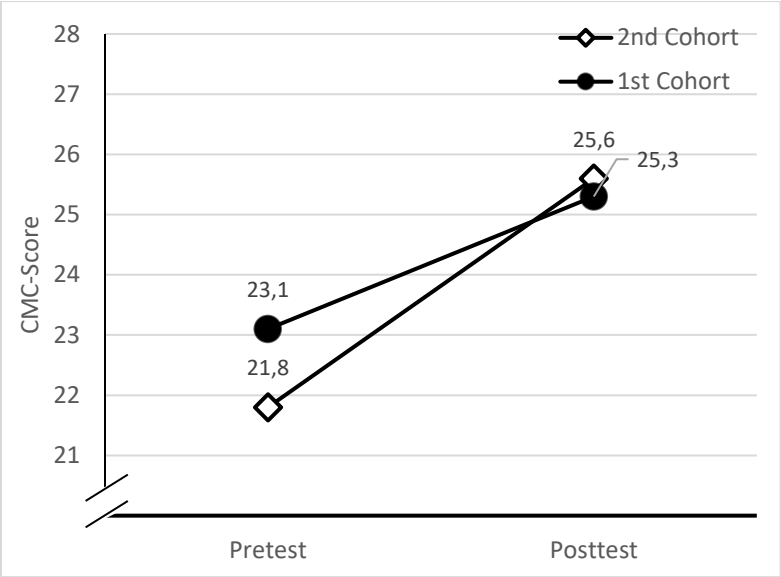


***Figure 11*. Mean Change between Pretest and Posttest for the 1st and 2nd cohort**

Thirdly, there was no significant first-order interaction effect of course-type, indicating that change in CMC-scores did not depend on instructional strategy ($F(1, 83) = 0.10$, $p = .75$, $\eta^2 = .001$), confirming hypothesis H2d (see Figure 12.).

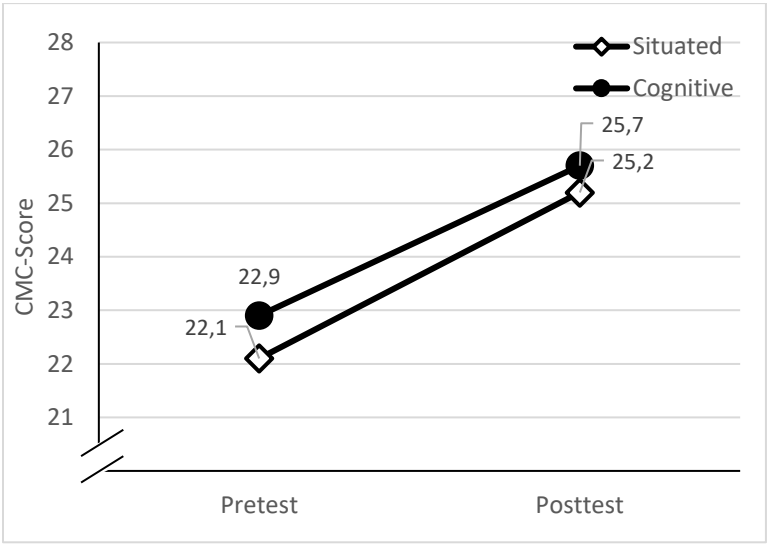


***Figure 12.* Mean Change between Pretest and Posttest for the Situated and Cognitive Instructional Strategies**

Finally, there were no significant second-order interaction effects between course-type and cohort, indicating that the effect of the instructional design did not depend on the degree of practical teaching experiences ($F(1, 83) = 0.27$, $p = .61$, $\eta^2 = .003$), rejecting hypotheses H2e and H2f (see Figure 13. and Figure 14.).

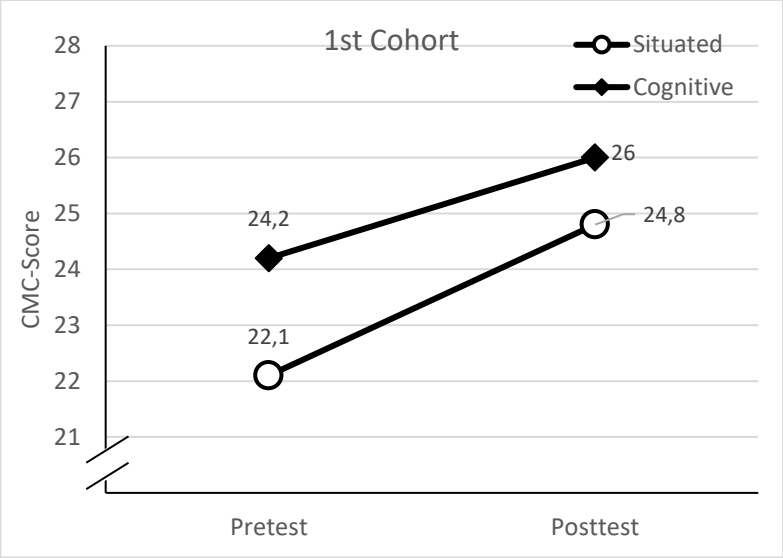


Figure 13. **Mean Change between Pretest and Posttest for the Situated and Cognitive Instructional Strategies of the first Cohort**

Figure 14. **Mean Change between Pretest and Posttest for the Situated and Cognitive Instructional Strategies of the second Cohort.**

8. Discussion

In this chapter, the results of the study are summarized and interpreted and their theoretical, methodological, and practical relevance is evaluated. Firstly, the hypotheses for statistical data analysis, formulated in chapter 6.7, are restated stated. On the basis of the results, presented in chapter 7, it is evaluated to what degree these hypotheses should be considered confirmed or rejected and how the results compare to other relevant research (8.1). Hypotheses are evaluated separately for construct validity (8.1.1) and course effectiveness (8.1.2). Secondly, the limitations of the current study are discussed (8.2). At first, limitations of the CMC-instrument and exploration of possible reasons for the values of test-reliability are discussed (8.2.1). Then, the sample and its representativeness (8.2.2) and limitations to the treatment (8.2.3) are discussed. Finally, the conclusions and relevance of these findings are evaluated and the sub questions answered (8.3). This is done by firstly discussing the model of situated competence of prospective teachers' competence (8.3.1). Secondly, the development of video tests for competence measurements is discussed (8.3.2). Finally, the use of videos for teaching classroom management competence in universities is discussed (8.3.3).

8.1 Main Results

8.1.1 Construct validity.

The hypothesis regarding the convergent validity of CMC was confirmed: Both video tests, CME (König, 2015b) and CMC, correlated significantly on both pretest and posttest (H1a). Moreover, as CMC was designed to match the content of the video-courses, a higher correlation between CMC and CME at pretest than at posttest is a further indication of CMC's construct validity. The low values of internal consistency and the medium correlation between pretest and posttest indicate that CMC's measurements include a relatively large portion of measurement error and therefore, should be considered imprecise. Consequently, CMCs reliability should be considered limited (see chapter 7.5). Unexpectedly, the values for internal consistency of CME (König, 2015b) were equally poor on both pretest and posttest, meaning that the portion of measurement error was equally high for CME. The lower the test-reliability, the lower the probability of finding

significant correlations. Nevertheless, as expected, significant correlations between CME and CMC were found at both pretest and posttest and convergent validity confirmed.

Hypotheses regarding discriminant validity were largely confirmed. As expected, CMC did not correlate to perceived self-efficacy on classroom management (H1b) (Pfitzner-Eden et al., 2014) and did not correlate to scores on general cognitive achievement (H1d) (Practical Semester Hesse, 2018). Moreover, CMC did not correlate to performance- (P) and continuity- (C) scores of FAIR-2 (Moosbrugger & Oehlschlägel, 2011) and quality scores (Q) of FAIR-2 did not correlate to CMC at posttest (H1c). However, Q-Scores did correlate significantly to CMC at pretest. This last result indicates that for the performance on CMC, a person's attentiveness did represent an advantage. This advantage, however, is only there when a low familiarity with the topic of classroom management exists. The fact that correlation with CMC only existed for Q-scores, indicates that the influence of attentiveness on CMC was limited to the attentiveness-aspect of careful scrutiny rather than to the aspect of quick scanning. The results regarding the correlation between test scores on CMC and scores on FAIR indicate that the influence of the trait of attentiveness on situation-specific skills on classroom management decreases with an increase in situation-specific skills. Apparently, if a test-participant was unfamiliar with the topic of classroom management, the ability to carefully scrutinize the teaching setting in the video clip was helpful for his or her performance on CMC. When a person increased his or her familiarity with classroom management through participation in the video-based courses, the advantage of attentiveness did not play a significant role anymore. Although the significant correlation between CMC and Q-scores on FAIR was not expected, it could be theoretically explained. Consequently, the above-discussed results indicate that construct validity was confirmed.

8.1.2 Course-effectiveness.

Results on self-efficacy

It was hypothesized that the cohort of prospective teachers that participated in the video-based course before starting with the practical semester would show higher degrees of perceived self-efficacy on classroom management than the cohort of prospective teachers that participated in the video-based courses after the practical semester. The t-test for independent samples,

however, showed a significant effect with a large effect-size in the opposite direction (see chapter 7.6), rejecting hypothesis H2a. Contrary to what was expected, these results show that prospective teachers that participated in the video-based courses rated their own ability to deal with classroom disruptions considerably lower than the prospective teachers who had not yet participated in the courses. The items measuring self-efficacy showed high values of internal consistency and were validated in previous studies (Pfitzner-Eden et al., 2014). Therefore, these measurements should be considered reliable.

These results are contrary to findings from similar research measuring the effect of video-based courses on self-efficacy (Gold et al., 2017; Kumschick et al., 2017). The measurement of self-efficacy of the groups in the studies of Gold et al. (Gold et al., 2017) and Kumschick et al. (Kumschick et al., 2017) was done in parallel, whereas the measurements in the current study were held at different times: assessment of the second cohort took place one year after the measurement of the first cohort, increasing the possibility of unknown intervening variables. Also, the study of Gold et al. (Gold et al., 2017) and Kumschick et al. (Kumschick et al., 2017) had a pre-post design, whereas the measurements of self-efficacy in the current study can be considered a cross-sectional measurement with waiting group design (see also Figure 1.). Additionally, the sample of Gold et al. (Gold et al., 2017) consisted of prospective students who, on average, were in their fourth term, with a relatively broad variability. The sample of Kumschick et al. (Kumschick et al., 2017) consisted only of master students. Therefore, it is assumed that most of the participating prospective teachers in these two studies had gathered their first teaching experience, either in an internship or working as a substitute teacher (Bäuerlein & Reintjes, 2018). As such, many of the prospective teachers in these studies based their self-efficacy on first practical experiences.

Contrarily, at the moment the prospective teachers of the current study rated their self-efficacy, hardly any teaching practice was available (see chapter 6.5 and Figure 5.). Prospective teachers of the second cohort had already participated in a video-based course, whereas prospective teachers from the first cohort participated in the course after completion of the practical semester. As such, prospective teachers from the second cohort had experienced a first reality-check through the analysis of the video-cases, whereas prospective teachers from the first cohort

rated their self-efficacy without reference to previous experiences.

It seems likely that the reality-check resulting from the video-based courses made a considerable number of prospective teachers from the second cohort to lower their perceived self-efficacy. It is therefore concluded that the substantial decrease in prospective teachers' self-efficacy on classroom management as a result of participation in the video-based courses is probably explained by their lack of practical experience. Relating these findings to the studies of Gold et al. (Gold et al., 2017) and Kumschick et al. (Kumschick et al., 2017), it is concluded that the participation in a video-based course only strengthens perceived self-efficacy when at pretest, prospective teachers base their perceived self-efficacy on personal practical teaching experience.

Results on situation-specific skills

Regarding the result of course-participation, several hypotheses were formulated. Firstly, it was expected that on average, the participants of all groups increased their situation-specific skills in classroom management, as measured by the score on CMC (H2b). Results of the mixed-design ANOVA (see chapter 7.6) show that a significant main effect was found with a large effect size, confirming hypothesis H2b. This means that on average, participants considerably increased their situation-specific skills in classroom management through participation in the video-based courses. Secondly, it was expected that prospective teachers' change in situation-specific skills would neither be dependent on the degree of previously acquired practical experience (H2c) nor on the course-type (H2d). Results of the analysis of variance showed no significant first-order interaction effects, confirming both hypotheses.

This means that the two course-types, the one following the cognitive strategy and the one following a situated strategy, can be considered equally effective. These results also indicate that the average effect of participation in both video-based courses did not depend on prospective teachers' previously gained practical experience. Moreover, two significant second-order interaction effects were expected. It was hypothesized that the first cohort which had previously participated in the practical semester would profit more from the course following the situated strategy (H2e), whereas the cohort that had not participated in the practical semester would profit more from the course following the cognitive strategy (H2f). Results of the analysis of

variance showed no significant second-order effects. Therefore, these hypotheses could not be confirmed, indicating that the effects of participation in the practical semester did not influence the effectiveness of either the situated or cognitive strategies of the courses.

Because the reported results are based on measurements of the video test CMC, limitations resulting from the measurement reliability discussed in chapter 8.1.1 apply and the results should be interpreted accordingly. This means, primarily, that measurements were imprecise. It was shown that not all assumptions for calculating a mixed-design ANOVA were met: For one of the four course-groups, the assumption of normally distributed residuals could not be confirmed at the posttest (see chapter 7.6). This means that an increased risk of type I or type II errors exists. Both facts, the limited reliability and the violation of the assumption of normally distributed residuals of one of eight measurements mean that the probability values of the significance tests should be interpreted with care. However, because the probability value for the main effect was smaller than $p < .001$ and the p-values of the two of the four interaction effects were $p = .61$ and $p = .75$, the risk of type I and type II errors for these interaction effects were assumed to be small. For the first-order interaction effect regarding the difference of the participation in the practical experience ($p = .20$), there is a possibility that a significant effect would have shown when a more reliable instrument would have been used. It is concluded that if first- or second-order effects occurred, they were not very big. In other words, in comparison to the strong average main effect of course-participation on situation-specific skills of prospective teachers, their previously acquired practical experience on the one hand and the instructional strategy of the courses on the other did not have a substantial impact.

In chapter 4.4, four studies were discussed that systematically compare the effects of instructional strategies of video-based courses on different aspects of prospective teachers' competence. These studies show several conceptual and methodological differences from the current study (see chapter 4.4.4). Also, none of these studies systematically controls the teaching experience of prospective teachers. As one of the four studies (Blomberg et al., 2014) differed both in the research setup and the dependent variable, the results cannot be compared to the current study. Below, the results of the current study are compared to the remaining three studies.

Syring et al. (Syring et al., 2015) showed that contrary to cognitive load theory, a problem-based course on classroom management did not produce higher degrees of extraneous cognitive load on prospective teachers and that no differences in motivation between the two groups existed. In the current study, the degrees of cognitive load of the different course-participants were not separately measured. It was shown, however, that both course-types caused a similar increase in situation-specific skills. This confirms the findings of Syring et al. (Syring et al., 2015) that a situated instructional approach does not necessarily lead to higher degrees of cognitive load than a cognitive instructional approach.

The study by Seidel et al. (Seidel et al., 2013) shows that for video-based courses, the more situated the competence aspect, the more favorable a course following a situated instructional strategy becomes. Prospective teachers participating in a video-based course based on cognitive strategy showed higher scores on declarative knowledge with a large effect size. This finding was confirmed for courses on classroom management by a study of Kumschick et al. (Kumschick et al., 2017). Seidel et al. (Seidel et al., 2013) also showed that a course based on the situated strategy increased scores on more situated competence aspect, like the quality of lesson-planning, with a large effect size. For the competence aspect of professional vision (see chapter 3.3), the course following a cognitive strategy showed higher scores with a medium effect size. As in the current study, no differences in effectiveness on situation-specific skills of the two instructional approaches could be found, the results of Seidel et al. (Seidel et al., 2013) could not be reproduced. A possible explanation for this lies in the fact that CMC proved to produce imprecise measurements. It is possible that the courses produced different effects that CMC was not able to depict (see chapter 7.5). However, because the results of the current study showed an almost identical change in CMC-scores with a low p-value, it seems unlikely that a more reliable instrument than CMC would have found a difference with a medium effect size. This indicates that different from what Seidel et al. (Seidel et al., 2013) conclude, a cognitive instructional strategy does not increase situation-specific skills or other situated competence aspects like professional vision, more than a situated instructional strategy.

8.2 Limitations of the study

There are several limitations of the current study that should be taken into account when interpreting the results. This section discusses the limitations regarding the instrument, the sample, and the treatment.

8.2.1 Instrument.

The strongest limitation of this research is related to the reliability of the CMC video test. As already mentioned, the low values of internal consistency and item-total correlation (see chapter 7.2.2), as well as the correlation between pretest and posttest (see chapter 7.5), indicate that CMC's reliability should be considered limited. CMC's limited reliability is reflected by the results of the correlation matrix and low values of the KMO test on sampling adequacy (see chapter 7.3). Additionally, it was shown that CME's values for internal consistency were even slightly lower than those for CMC (see chapter 7.4.2).

There are typically three reasons for low values of test reliability (Bühner, 2011). A first factor, which possibly lowers values of reliability, is limited variance. However, as item-analysis showed mean-variance for the 20 items of CMC lies at $M_{Var} = 0.53$, it is not expected that item-variance plays a considerable role. Secondly, low values of reliability can be related to random error, for example, through the effect of tiredness or external influences during test-administration (Bühner, 2011). Because the conditions of testing were identical between the different measurements at pretest and between pretest and posttest, it is unlikely that external influences played a big role. At pretest and posttest, CMC and CME were administered directly after each other with no breaks lasting around 75 minutes in total. Although this represents a considerable time, most test-participants were not concentrated on the tests for the entire period. Between video clips, participants needed to wait for all other participants to finish before showing the next video clip. However, the lower values for internal consistency on CME could be explained by the fact that it was administered after test-participants already took the CMC video test. Another potential source of random error is the quality of coding. However, as it was shown that values of several indexes of inter-rater reliability were good for both CME and CMC (see chapter 7.1), the quality of coding is not expected to play a considerable role regarding random error. It is

concluded that although tiredness could play a minor role, it is unlikely that low reliability is explained to a considerable degree by random error like tiredness or coding quality.

A third factor that could possibly be responsible for low values or test-reliability is the heterogeneity of its items (Bühner, 2011). Homogeneity refers to the degree of inter-correlation between items, as indicated by the value of item-total correlation. High degrees of homogeneity lead to high degrees of internal consistency, which is an indication of a test's reliability. As factor analysis showed low intercorrelations between the items, it can be inferred that CMC contains a relatively low degree of homogeneity. These findings indicate that the reason for the CMC's limited reliability probably lies within the formulation of its items, as the items firstly represent a heterogeneous construct and secondly insufficiently reflect the theoretical constructs. The following section discusses some of the issues relating to the items.

CMC's items were validated through three rounds of expert-ratings (see chapter 6.2.6). The second round of expert-rating rated criteria on average as $M = 4.0$ on a five-point Likert-scale. The results of this rating led to the exclusion or reformulation of several criteria that the experts rated as inappropriate for use in the video test. Also, a pilot assessment was held (see chapter 6.2.4). In this pilot assessment, among other things, test-participants were asked to rate the instrument. Although users rated the items' comprehensibility as mediocre, the test as a whole, both items and videos were rated as appropriate within the pilot sample. Moreover, after the elimination of some items with obvious inadequate psychometric values, a core of 18 good items was left with an acceptable to good value of internal consistency (see chapter 6.2.4). These results were interpreted to indicate that CMC was suitable for the assessment of classroom management competence for prospective teachers.

Another indication that the items represented a source of error is shown by the fact that a total of eight items showed medium to good values for item-total correlation at pretest, but poor values at posttest or the other way around. This indicates that several items are mostly suitable for measuring CMC at pretest when no familiarity with the construct exists and other items that are only suitable when test-participants are more familiar with the construct of classroom management. It seems probable that the subjective interpretation of the items differs depending

on these conditions.

In sum, the above-discussed results indicate that CMC could have been more extensively piloted with prospective teachers. As a piloting study and expert-validation already took place, a think-aloud study (Pepper, Hodgen, Lamesoo, Kõiv, & Tolboom, 2016) with a smaller panel of prospective teachers could have provided additional information that provides insight into further reasons for low reliability. Among other things, this could provide more information on why some items perform better at pretest and other items better at posttest.

A final instrument-related limitation of this current study is the fact that due to constraints of resources, it could not be assessed to what degree test practice effects occurred (Hausknecht, Halpert, Di Paolo, & Moriarty Gerrard, 2007). Although the possibility of practice effects cannot be excluded, a considerable test-effect is not expected, because no effects were found in a similar study in which CME is applied in a control group (Kramer et al., 2017).

8.2.1 Sample and generalizability.

Because the sample consisted of prospective teachers who voluntarily registered for the course, the sample should not be regarded as representative. Results are therefore limited to the sample and cannot be generalized to the population of prospective teachers. Moreover, the sample size was limited to the number of prospective teachers that voluntarily registered to the course and can be considered small. A small sample size is prone to outliers and violations of normally distributed values and residuals (Field, 2013). Although violations to the assumptions were minimal and therefore their risk of influence on the p-values can be considered low (see chapter 7.6), a larger sample size with the four groups would have considerably increased the possibility of normally distributed values and residuals and would, therefore, have improved the correct representation of the *p*-values.

Another limitation related to the sample is the considerable degree of sample attrition. This sample attrition consisted of 10 prospective teachers that participated in pretest, but did not show up for the courses and three additional prospective teachers of which the pseudonyms of posttest could not be matched to the pseudonyms of the pretest. Of these 13 prospective

teachers, it cannot be determined to what degree their attrition was random or represents a systematical bias.

8.2.3 Treatment.

Analysis of the subjective experience of prospective teachers showed a considerable difference between the ratings of the prospective teachers of the first cohort and the prospective teachers of the second cohort (see chapter 7.4.5 and Table 20). As discussed in chapter 8.1.2, there is a possibility that these differences are the result of practice effects of the teachers of the course. Because of time constraints, the course-designs could not be piloted. As such, experiences from teaching the first cohort could have influenced the teaching of the second cohort. However, because of the standardization of lectures, materials and exercises, these effects are expected to be relatively small.

Because of the limitations of the content of the available video clips, two deviations from the guidelines regarding the systematic differences between the two course-types were made (see chapter 6.3.3). It was expected that the depicted situations in the video clips were not focused enough for participants of the course with the situated instructional design to initiate a discussion on the respective strategy on their own accord. The decisions for these deviations followed on extensive discussions between the teachers. Although measures were taken to minimize the impact of these exceptions, they represent a deviation from the systematic difference between the two courses.

8.3 Conclusions

In chapter 1, it was stated that the current study aims to make contributions on three levels: Firstly, on a theoretical level, the study aims to advance the discussion on the options for situated models of prospective teachers' competence. Secondly, on a methodological level, the results aim to contribute to the possibilities and conditions for the development of video tests. Finally, on the practical level, the results intend to contribute to the research regarding the possibilities of video-based courses on classroom management for prospective teachers. In the following sections, for each of these levels, the respective research questions are answered, the

contributions of the results of the current study are evaluated, and the possibilities for future research are formulated.

8.3.1 The PID-model and situation-specific skills.

Regarding the concept of competence, the theoretical discussion in chapter three showed that the PID-model provides a promising model of situated teachers' competence (see chapter 3.4). This model seems especially relevant for the aspect of classroom management (see chapter 3.7). The PID-model focuses on three situation-specific skills: perception, interpretation, and decision-making (Blömeke, Gustafsson, & Shavelson, 2015a; Stahnke et al., 2016). As discussed in chapter 3.6, there are two known video tests measuring a situated form of classroom management competence: CME (König, 2015b) and VIU (Gold & Holodynski, 2015). However, confirmatory factor analysis on CME showed that the underlying dimensions of CME, *accuracy of perception*, *holistic perception* and *justification of action*, could not be empirically separated and CME should, therefore, be considered as a unidimensional construct (König, 2015b). A confirmatory factor analysis on the data of two validation studies within project VIU (Gold & Holodynski, 2015) empirically confirmed three underlying dimensions regarding classroom management: *monitoring*, *managing momentum* and *rules and routines*. However, the dimensions of *noticing* and *interpretation* that were also used for item-formulation, could not be confirmed by the data.

EFA performed within the current study showed that also the CMC video test could not empirically separate between dimensions of noticing and interpretation (see chapter 7.3). As such, the current research confirms that there is reason to doubt that perception and interpretation represent different aspects of competence of prospective teachers on classroom management.

Research on attentiveness indicates that this personal trait could be relevant for performance on video tests (Copeland 1987). Results from the current study found no substantial evidence that attentiveness influenced test-scores of CMC (see chapter 7.5). However, because of CMC's limited reliability (see chapter 8.2.1) and the fact that Q-scores of FAIR-2 were significantly related to CMC's score at pretest, the possibility of an influence of attentiveness on performance on video tests cannot be ruled out.

The results of the current study indicate that changes in the interrelation between different competence aspects occurred as a result of the video-based courses: where situation-specific skills increased, self-efficacy decreased (see chapter 7.6 and chapter 8.1.3). Empirical research confirms that for prospective teachers, self-efficacy is related to classroom management competence (Gold et al., 2017; Kumschick et al., 2017). However, no studies are known that support the thesis that self-efficacy is reduced due to familiarity with the topic of classroom management.

It is concluded that although the PID-model is promising as a model for the situated assessment of teachers' competence, further research is needed to test some of its fundamental theoretical assumptions. Especially the assumptions regarding the difference between perception and interpretation, the assumed positive relation between situation-specific skills and beliefs, and the assumption of predictive validity for teaching performance are in need of further empirical testing. Furthermore, further research could further go into the question to what degree attentiveness plays a role in the performance in video tests. Lastly, further research could focus on investigating the role of self-efficacy within classroom management competence of prospective teachers. More precisely, the specific hypothesis could be tested, to what degree experience or familiarity with the topic of classroom management plays a role in values of perceived self-efficacy.

8.3.2 Video tests for competence measurement.

Chapter three discussed two examples of video tests that measure classroom management competence (see chapter 3.5). These video tests are illustrative of the fact that the method of using video tests for the measurement of situated competence is gaining popularity (Seidel & Thiel, 2017). Video tests are used in different forms: Items can be formatted as rating-items, open-response, or multiple-choice items (Lindmeier, 2013). However, the used items, their operationalization, formulation and coding are often not wholly reported. Although from a pragmatic perspective, the reasons might be valid, this omission hinders the exact reproduction and comparison of the different instruments (Neuweg, 2015).

The current study aims to provide readers with as much transparency as possible regarding the

design process of the video test. Some of the possible reasons for this were discussed in chapter 8.2.1. However, because results between the pilot study and the main study deviated so strongly, no apparent single reason for the unsatisfactory values could be identified. This indicates that the process of development of a video test includes several pitfalls and should not be underestimated. Future video tests should, therefore, be extensively piloted, preferable through think-aloud studies (Pepper et al., 2016).

Additionally, the range of different types of video tests is extensive. This variety includes the use, length, content, and complexity of the video-clips and the item-type and test-setup. First initiatives have compared different types of video tests and linking the design and measured effects (Gold et al., 2016). Future research could build on these results and focus on comparing the results of different video tests on the same sample. More specifically, specific test-characteristics like the length and complexity of the video clips, item-format and complexity could be systematically varied to investigate the effects on the participants' test-scores, sense of immersion, cognitive load, and motivation.

8.3.3 Teaching classroom management through video-based case-analysis.

For the learning of classroom management competence for prospective teachers, chapter four discusses two relevant opportunities to learn: internships (see chapter 4.2) and video-based courses (see chapter 4.3). It was shown that on average, internships provide more realistic notions on teaching and the teacher profession (Hascher, 2012b), and increase general pedagogical knowledge (König, 2013), self-efficacy (Gröschner & Schmidt, 2012), and professional vision (Mertens & Gräsel, 2017). As no research was found on the effects of internships on classroom management competence, the question arises, to what degree these changes on a general level also occur for the specific competence aspect of classroom management.

The current study did not systematically evaluate the effects of the practical semester. However, as the main difference between the two cohorts lies in their participation in the practical semester, differences of the scores on CMC between the two cohorts at pretest are assumed to be the result of differences in practical experience. However, the current study showed no

significant difference between the two cohorts. Nevertheless, because of the limited reliability of CMC on the one hand and the result of the t-test, $t(85) = 1.16$, $p = .25$ (see chapter 7.7.2), it is probable, that a more precise instrument would have found significant differences at pretest. The fact that prospective teachers from the first cohort rated the knowledge gains as considerably smaller and the pace of the course as lower represent further indications that a relevant difference regarding classroom management between the two cohorts existed. As the current study only used a video test, it could not be further investigated what these differences consisted of. However, the results of the current study indicate that for the conveying of classroom management competence, differences in practical experience are relevant. Courses aiming at conveying of classroom management competence are therefore encouraged to explicitly relate to the practical experience of their learners.

Courses based on case-analysis of videos are heavily based on the availability of appropriate video clips (Blomberg et al., 2013; Wirtz, 2002). Especially for the topic of classroom management, which involves teacher's dealing with (potential) problematic student behavior, the number of adequate available video clips is limited. Therefore, it proved to be challenging to find appropriate video clips (see chapter 6.3.2). This was especially a challenge for the courses following a situated instructional strategy, because following the logic of this instructional strategy, the teacher's aim was to minimize the structuring of learners' discussions. This could only be solved by deviating from guidelines of the instructional design on two occasions (see chapter 6.3.3 and chapter 8.2.3). Researchers attempting a similar comparison of instructional approach as the current study should therefore be aware that the quality and content of the available video clips is even more crucial to courses following a situated approach. Accordingly, great care and sufficient time with the selection of video clips should be taken.

A promising alternative for teachers to guide learners' attention without providing theory, lies in developing of staged video clips. Staged videos give teachers full control over the designing process and its portrayed teaching (Seidel & Thiel, 2017). For the topic of classroom management, first projects for the development of such cases have been initiated (Barth, 2016; Bönte, Nissing, Lenske, Dicke, & Leutner, 2017; Piwowar et al., 2017).

Regarding the learning from video clips, past research shows that video-based case-analysis represents an effective way to convey situation-specific skills on classroom management (see chapter 4.3). However, theory indicates that the complexity of video clips can represent a burden on cognitive load, potentially limiting the effectiveness of learning processes of novice teachers (see chapter 4.4). Although not systematically measured, the current study found no evidence for the claim that cognitive load plays a strong role, as learning effects between the two instructional designs were identical. For the teaching of classroom management competence at university, this means that both instructional approaches represent a valid method. This should be seen as an argument to combine the strengths of both instructional approaches and offer courses in which both approaches are used (Seidel et al., 2013).

To further investigate the influence of teaching experience and instructional design on the effectiveness of prospective teachers' courses on classroom management, future research could focus on the following points: Firstly, the influence of internships on beliefs on classroom management, like for example on classroom control (Martin & Yin, 2006) could be investigated. This would provide relevant information for teacher education, for example with the clearing of misconceptions regarding issues of discipline. Secondly, as the current study indicates that participation in an internship influences classroom management competence, further research could focus on specifically evaluating the influence of participation in an internship on situation-specific skills on classroom management in a pretest-posttest experimental setup. Thirdly, although more complex in its design, further research could focus on longitudinal studies that pose the question whether a video-based course on classroom management should be visited as a preparation of the internship, or that alternatively practical experience is needed to fully profit from the video-based course. Although the current research does not give reason to believe that a strong difference exists, it is relevant to programs of teacher education to have more substantial evidence regarding this question.

Fourthly and lastly, the discussion to what degree cognitive load has an impairing factor on the ability to analyze video clips is of high relevance for teacher educators. The current study, as well as previous research (Syring et al., 2015), indicates that cognitive load does not play a significant role regarding instructional strategy. These results represent deviations from what would be

expected from theory. As such, it could be further explored to what degree cognitive load plays a relevant role when analyzing video clips. Therefore, research systematically controlling other characteristics of videos, like for example its length and complexity or the comparison of own teaching against unknown teaching would provide highly relevant information for educations that use of video clips within teacher education.

REFERENCES

Abramson, J. (2016). WinPepi. Retrieved from http://www.brixtonhealth.com/pepi4windows.html

Anderson, J., Greeno, J., Reder, L. M., & Simon, H. A. (2000). Perspectives on Learning, Thinking, and Activity. *Educational Researcher*, *29*(4), 11–13.

Anderson, J., Reder, L. M., & Herbert, S. A. (1996). Situated Learning and Education. *Educational Researcher*, *25*(4), 5–11.

Anderson, L., Evertson, C. M., & Emmer, E. T. (1980). Dimensions in Classroom Management Derived from Recent Research. *Journal of Curriculum Studies*, *12*(4), 343–356.

Anselm, S. (2010). "Wissen ist gut, doch Können ist besser.". In J. Abel (Ed.), *Wirkt Lehrerbildung?: Antworten aus der empirischen Forschung* (pp. 155–163). Münster: Waxmann.

Appel, J. (2016). Zusammenhänge zwischen Aufgaben und Schülerbeteiligungsverhalten im Unterrichtsprozess: Ergebnisse einer Videostudie. In S. Keller & C. Reintjes (Ed.), *Aufgaben als Schlüssel zur Kompetenz: Didaktische Herausforderungen, wissenschaftliche Zugänge und empirische Befunde* (1st ed., pp. 261–272). Münster: Waxmann.

Appel, J., & Rauin, U. (2016). Quantitative Analyseverfahren in der videobasierten Unterrichtsforschung. In U. Rauin, M. Herrle, & T. Engartner (Ed.), *Grundlagentexte Methoden. Videoanalysen in der Unterrichtsforschung: Methodische Vorgehensweisen und Anwendungsbeispiele* (pp. 130–172). Weinheim: Beltz Juventa.

Bandura, A. (1994). Self-Efficacy. In V. S. Ramachaudran (Ed.), *Encyclopedia of Human Behavior* (4th ed., pp. 71–81). New York: Academic Press.

Bandura, A. (1997). *Self-efficacy: The Exercise of Control*. New York: W.H. Freeman.

Barth, V. L. (2016). *Professionelle Wahrnehmung von Störungen im Unterricht*. [Place of publication not identified]: Springer Science and Business Media; Springer VS.

Bäuerlein, K., & Reintjes, C. (2018). *Studienunabhängige Selbstprofessionalisierung in der Schule?* Annual Convention of the Gesellschaft für Empirische Bildungsforschung, Basel.

Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift Für Erziehungswissenschaft*, *9*(4), 469–520.

Baumert, J., & Kunter, M. (2011). Das Kompetenzmodel von COACTIV. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Ed.), *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV* (pp. 29–53). Münster, New York, NY, München, Berlin: Waxmann.

Baumert, J., & Kunter, M. (2013). The COACTIV Model of Teachers' Professional Competence. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Ed.), *Cognitive Activation in the Mathematics Classroom and Professional Competence of Teachers* (pp. 25–48). Boston, MA: Springer US.

Baumert, J., Kunter, M., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (2011). Professionelle Kompetenz von Lehrkräften, kognitiv aktivierender Unterricht und die mathematische Kompetenz von Schülerinnen und Schülern (COACTIV): Ein Forschungsprogramm. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Ed.), *Professionelle Kompetenz von Lehrkräften:*

Ergebnisse des Forschungsprogramms COACTIV (pp. 7–25). Münster, New York, NY, München, Berlin: Waxmann.

Berliner, D. (1992). The Nature of Expertise in Teaching. In F. Oser, A. Dick, & J. L. Patry (Ed.), *Effective and responsible teaching* (pp. 227–248). San Francisco: Jassey-Bass.

Biaggi, S., Krammer, K., & Hugener, I. (2013). Vorgehen zur Förderung der Analysekompetenz in der Lehrerbildung mit Hilfe von Unterrichtsvideos: Erfahrungen aus dem ersten Schuljahr. *Seminar*, *2*, 26–34.

Bischoff, S., Brühwiler, C., & Baer, M. (2005). Videotest zur Erfassung adaptiver Lehrkompetenz. *Beiträge Zur Lehrerinnen- Und Lehrerbildung*, *23*(3), 382–397.

Blomberg, G., Sherin, M. 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.

Blomberg, G. A., Renkl, A., Sherin, G. M., Borko, H., & Seidel, T. (2013). Five research-based heuristics for using video in pre-service teacher education. *Journal for Educational Research Online*, *5*(1), 90–114.

Blömeke, S. (Ed.). (2008). *Professionelle Kompetenz angehender Lehrerinnen und Lehrer: Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematikstudierende und -referendare; erste Ergebnisse zur Wirksamkeit der Lehrerausbildung*. Münster, New York, München, Berlin: Waxmann.

Blömeke, S., Busse, A., Kaiser, G., König, J., & Suhl, U. (2016). The relation between content-specific and general teacher knowledge and skills. *Teaching and Teacher Education*, *56*, 35–46.

Blömeke, S., Gustafsson, J. E., & Shavelson, R. (2015a). Approaches to Competence Measurement in Higher Education. *Zeitschrift Für Psychologie*, *223*(1), 1–2.

Blömeke, S., Gustafsson, J. E., & Shavelson, R. (2015b). Beyond Dichotomies. *Zeitschrift Für Psychologie*, *223*(1), 3–13.

Blömeke, S., Kaiser, G., & Lehmann, R. (Ed.). (2010). *TEDS-M 2008: Professionelle Kompetenz und Lehrgelegenheiten angehender Mathematiklehrkräfte für die Sekundarstufe I im internationalen Vergleich*. Münster: Waxmann.

Blömeke, S., & König, J. (2010). Messung des pädagogischen Wissens: Theoretischer Rahmen und Teststruktur. In S. Blömeke, G. Kaiser, & R. Lehmann (Ed.), *TEDS-M 2008: Professionelle Kompetenz und Lehrgelegenheiten angehender Mathematiklehrkräfte für die Sekundarstufe I im internationalen Vergleich* (pp. 239–263). Münster: Waxmann.

Blömeke, S., König, J., Suhl, U., Hoth, J., & Döhrmann, M. (2015). Wie Situationsbezogen ist die Kompetenz von Lehrkräften: Zur Generalisierbarkeit der Ergebnisse von videobasierten Performanztests. *Zeitschrift Für Pädagogik*, *61*(3), 310-327.

Blömeke, S., Suhl, U., & Döhrmann, M. (2012). Zusammenfügen was zusammengehört: Kompetenzprofile am Ende der Lehrerausbildung im internationalen Vergleich. *Zeitschrift Für Pädagogik*, *58*(4), 422–440.

Blüthmann, I., Ophardt, D., Thiel, F., & Felsberger, G. (2011). Wissenserwerb zum Thema Klassenmanagement im Lehramtsstudium: Individuelle und studienbezogene Einflussfaktoren. *Unterrichtswissenschaft*, *39*(4), 290–309.

Bönte, J., Nissing, R., Lenske, G., Dicke, T., & Leutner, D. (2017). The impact of staged video vignettes on teachers' nonverbal classroom management knowledge.: Poster presentation on the 21th

Conference of the Junior Researchers of EARLI (JURE), Tampere. Retrieved from https://www.uni-due.de/biwi/llp/de/vita_boente

Borich, G. D. (2011). *Observation skills for effective teaching* (6th ed). Boston, MA: Allyn & Bacon/Pearson.

Borich, G. D. (2014). *Effective Teaching Methods: Research-based Practice*. Boston: Pearson.

Borko, H. (2016). Methodological contributions to video-based studies of classroom teaching and learning: a commentary. *ZDM*, *48*(1-2), 213–218.

Bromme, R. (1992). *Der Lehrer als Experte: Zur Psychologie des professionellen Wissens*. Bern: Verlag Hans Huber.

Bromme, R. (2001). Teacher Expertise. In F. E. Weinert (Ed.), *International Encyclopedia of the Behavioral Sciences* (p. 15469). London: Pergamon.

Brophy, J. E. (2006). History of Research on Classroom Management. In C. M. Evertson & C. S. Weinstein (Ed.), *Handbook of classroom management: Research, practice, and contemporary issues* (Ch. 2, pp. 17–43). New York, N.J.: Routledge.

Brouwers, A., & Tomic, W. (2000). A longitudinal study of teacher burnout and perceived self-efficacy in classroom management. *Teaching and Teacher Education*, *16*(2), 239–253.

Bruckmaier, G., Krauss, S., Blum, W., & Leiss, D. (2016). Measuring mathematics teachers' professional competence by using video clips (COACTIV video). *ZDM*, *48*(1-2), 111–124.

Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* (3rd Ed). *PS Psychologie*. München: Pearson Studium.

Burns, R. B., & Anderson, L. (1987). The Activity Structure of Lesson Segments. *Curriculum Inquiry*, *17*(1), 31–53.

Canter, L. (2010). *Lee Canter's assertive discipline: Positive behavior management for today's classroom* (4th ed., new & rev). Bloomington IN: Solution Tree Press.

Carter, K., Cushing, K., Sabers, D., Stein, P., & Berliner, D. (1988). Expert-Novice Differences in Perceiving and Processing Visual Classroom Information. *Journal of Teacher Education*, *39*(3), 25–31.

Casale, G., Strauß, S., Hennemann, T., & König, J. (2016). Wie lässt sich Klassenführungsexpertise messen?: Überprüfung eines videobasierten Erhebungsinstruments für Lehrkräfte unter Anwendung der Generalisierbarkeitstheorie. *Empirische Sonderpädagogik*. (2), 119–139.

Cicchetti, D. V., & Feinstein, A. R. (1990). High agreement but low kappa: II. Resolving the paradoxes. *Journal of Clinical Epidemology*, *43*(6), 551–558.

Cocard, Y., & Krähenbühl, S. (2013). *Pädagogisches Wissen in Klassenführung: Forschungsbericht*. Bern.

Cocard, Y., & Krähenbühl, S. (2015). Pädagogisches Wissen in Klassenführungskonzepten angehender Lehrpersonen. *Lehrerbildung Auf Dem Prüfstand*, *8*(1), 55–76.

Cochran-Smith, M., Feiman-Nemser, S., & McIntyre, D. J. (Ed.). (2008). *Handbook of research on teacher education: Enduring questions in changing contexts* (3rd ed.). New York, London: Routledge. Taylor & Francis Group.

Cochran-Smith, M., & Zeichner, K. (Ed.). (2005). *Studying teacher education: The report of the AERA Panel on Research and Teacher Education*. Mahwah, N.J: Lawrence Erlbaum Associates.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, N.J.: L. Erlbaum Associates.

Combe, A., & Kolbe, F.-U. (2008). Lehrerprofessionalität: Wissen, Können, Handeln. In W. Helsper (Ed.), *Handbuch der Schulforschung* (2nd ed., pp. 857–875). Wiesbaden: VS, Verlag für Sozialwissenschaften.

Copeland, W. D. (1987). Classroom Management and Student Teachers' Cognitive Abilities: A Relationship. *American Educational Research Journal*, *24*(2), 219–236.

Copeland, W. D., Birmingham, C., DeMeulle, L., D'Emidio-Caston, M., & Natal, D. (1994). Making Meaning in Classrooms: An Investigation of Cognitive Processes in Aspiring Teachers, Experienced Teachers and Their Peers. *American Educational Research Journal*, *31*(1), 166–196.

Darling-Hammond, L. (1999). The Case for University-Based Teacher Education. In R. A. Roth (Ed.), *The role of the university in the preparation of teachers* (pp. 333–346). London: Falmer Press.

Dicke, T., Parker, P. D., Marsh, H. W., Kunter, M., Schmeck, A., & Leutner, D. (2014). Self-efficacy in classroom management, classroom disturbances, and emotional exhaustion: A moderated mediation analysis of teacher candidates. *Journal of Educational Psychology*, *106*(2), 569–583.

Dinkelaker, J., & Herrle, M. (2009). *Erziehungswissenschaftliche Videographie: Eine Einführung* (1st Ed). *Qualitative Sozialforschung*. Wiesbaden: VS, Verlag für Sozialwissenschaften.

Döring, N., & Bortz, J. (2016a). Forschungs- und Wissenschaftsethik. In N. Döring & J. Bortz (Ed.), *Springer-Lehrbuch. Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5th ed., pp. 121–139). Berlin: Springer.

Döring, N., & Bortz, J. (2016b). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5th Ed). *Springer-Lehrbuch*. Berlin: Springer.

Doyle, W. (1985). Recent Research on Classroom Management: Implications for Teacher Education. *Journal of Teacher Education*, *36*(3), 31–35.

Doyle, W. (1984). How Order is Achieved in Classrooms: an Interim Report. *Journal of Curriculum Studies*, *16*(3), 259–277.

Doyle, W. (1986). Classroom Organization and Management. In M. C. Wittrock (Ed.), *Handbook of Research on Teaching* (3rd ed.). New York: Macmillan.

Doyle, W. (2006). Ecological Approaches to Classroom Management. In C. M. Evertson & C. S. Weinstein (Ed.), *Handbook of classroom management: Research, practice, and contemporary issues*, (pp. 97–125). New York, N.J.: Routledge.

Duffy, G. G. (1990). Whatever became of Teacher Effectiveness: Editorial Comment. *Reading Psychology*, *11*(4), 3–10.

Dunekacke, S., Jenßen, L., Eilerts, K., & Blömeke, S. (2016). Epistemological beliefs of prospective preschool teachers and their relation to knowledge, perception, and planning abilities in the field of mathematics: A process model. *ZDM*, *48*(1-2), 125–137.

Eid, M., Gollwitzer, M., & Schmitt, M. (2015). *Statistik und Forschungsmethoden* (4th Ed.). Weinheim: Beltz.

Emmer, E. T., Evertson, C. M., & Anderson, Linda. (1980). Effective Classroom Management at the Beginning of the School Year. *The Elementary School Journal*, *80*(5), 219–231.

Emmer, E. T., & Stough, L. M. (2001). Classroom Management: A Critical Part of Educational Psychology, With Implications for Teacher Education. *Educational Psychologist*, *36*(2), 103–112.

Erickson, F. (2011a). On Noticing Teacher Noticing. In M. G. Sherin, V. R. Jacobs, & R. A. Philipp (Ed.), *Studies in mathematical thinking and learning. Mathematics Teacher Noticing: Seeing through Teachers' Eyes* (pp. 17–34). London: Routledge.

Erickson, F. (2011b). Uses of video in social research: A brief history. *International Journal of Social Research Methodology*, *14*(3), 179–189.

Erickson, F., Boersma, D., Brown, M., Kirschner, B., Lazarus, B., Pelissier, C., & Thomas, D. (1986). *Teachers' practical ways of seeing and making sense: A final report*. Washington, D.C.: Office of Educational Research and Improvement.

Evertson, C. M., & Emmer, E. T. (2009). *Classroom Management for Elementary Teachers*. New Jersey: Pearson Education.

Evertson, C. M., & Weinstein, C. S. (2006a). Classroom Management as a Field of Inquiry. In C. M. Evertson & C. S. Weinstein (Ed.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 3–15). New York, N.J.: Routledge.

Evertson, C. M., & Weinstein, C. S. (Ed.). (2006b). *Handbook of classroom management: Research, practice, and contemporary issues*. New York, N.J.: Routledge.

Feinstein, A. R., & Cicchetti, D. V. (1990). High agreement but low Kappa: I. the problems of two paradoxes. *Journal of Clinical Epidemiology*, *43*(6), 543–549.

Field, A. P. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). London: Sage.

Fishman, J. A., & Galguera, T. (2003). *Introduction to test construction in the social and behavioral sciences: A practical guide*. Lanham, MD: Rowman & Littlefield Publishers.

Fisseni, H. J. (2004). *Lehrbuch der psychologischen Diagnostik: Mit Hinweisen zur Intervention* (3rd Ed.). Göttingen: Hogrefe.

Friedman, I. A. (2006). Classroom Management and Teacher Stress and Burnout. In C. M. Evertson & C. S. Weinstein (Ed.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 925–944). New York, N.J.: Routledge.

Gaudin, C., & Chaliès, S. (2015). Video viewing in teacher education and professional development: A literature review. *Educational Research Review*, *16*, 41–67.

Gettinger, M., & Kohler, K. M. (2006). Process-Outcome Approaches to Classroom Mananagement and Effective Teaching. In C. M. Evertson & C. S. Weinstein (Ed.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 73–95). New York, N.J.: Routledge.

Goethe University Frankfurt. Studierendenstatistik. Retrieved from http://www.uni-frankfurt.de/52565158/studierendenstatistik_pro_semester?legacy_request=1

Goeze, A. (2010). Was ist ein guter Fall?: Kriterien für die Entwicklung und Auswahl von Fällen für den Einsatz in der Aus- und Weiterbildung. In J. Schrader (Ed.), *EB-Buch: Vol. 31. Mediengestützte Fallarbeit: Konzepte, Erfahrungen und Befunde zur Kompetenzentwicklung von Erwachsenenbildnern* (pp. 125–145). Bielefeld: Bertelsmann.

Goeze, A., Hetfleisch, P., & Schrader, J. (2013). Wirkungen des Lernens mit Videofällen bei Lehrkräften. *Zeitschrift Für Erziehungswissenschaft*, *16*(1), 79–113.

Gold, B., Förster, S., & Holodynski, M. (2013). Evaluation eines videobasierten Trainingsseminars zur Förderung der professionellen Wahrnehmung von Klassenführung im Grundschulunterricht. *Zeitschrift Für Pädagogische Psychologie*, *27*(3), 141–155.

Gold, B., Hellermann, C., & Holodynski, M. (2016). Professionelle Wahrnehmung von Klassenführung: Vergleich von zwei videobasierten Erfassungsmethoden. In D. Prinz & K. Schwippert (Ed.), *Der Forschung, Der Lehre, Der Bildung: Aktuelle Entwicklungen der Empirischen Bildungsforschung* (pp. 103–118). Münster, New York: Waxmann.

Gold, B., Hellermann, C., & Holodynski, M. (2017). Effekte videobasierter Trainings zur Förderung der Selbstwirksamkeitsüberzeugungen über Klassenführung im Grundschulunterricht. *Zeitschrift Für Erziehungswissenschaft*, *20*(S1), 115–136.

Gold, B., & Holodynski, M. (2015). Development and Construct Validation of a Situational Judgment Test of Strategic Knowledge of Classroom Management in Elementary Schools. *Educational Assessment*, *20*(3), 226–248.

Gold, B., & Holodynski, M. (2017). Using Digital Video to Measure the Professional Vision of Elementary Classroom Management: Test Validation and Methodological Challenges. *Computers & Education*, *107*, 13–30.

Good, T. L., & Brophy, J. E. (2008). *Looking in Classrooms* (10th Edition). Boston: Pearson Education.

Goodwin, C. (1994). Professional Vision. *American Anthropologist*, *96*(3), 606–633.

Goodwin, C. (2001). Practices of Seeing: Visual Analysis: An Ethnomethodological Approach. In T. van Leeuwen & C. Jewitt (Ed.), *Handbook of visual analysis* (pp. 157–182). London, Thousand Oaks [Calif.]: Sage.

Greeno, J. (1998). The Situativity of Knowing, Learning and Research. *American Psychologist*, *53*(1), 5–26.

Gröschner, A., Müller, K., Bauer, J., Seidel, T., Prenzel, M., Kauper, T., & Möller, J. (2015). Praxisphasen in der Lehrerausbildung – Eine Strukturanalyse am Beispiel des gymnasialen Lehramtsstudiums in Deutschland. *Zeitschrift Für Erziehungswissenschaft*, *18*(4), 639–665.

Gröschner, A., & Schmidt, C. (2012). Kompetenzentwicklung im Praktikum?: Entwicklung eines Instruments zur Erfassung von Kompetenzeinschätzungen und Ergebnisse einer Befragung von Lehramtsstudierende im betreuten Blockpraktikum. *Lehrerbildung Auf Dem Prüfstand*, *5*(2), 112–128.

Gröschner, A., Schmitt, C., & Seidel, T. (2013). Veränderung subjektiver Kompetenzeinschätzungen von Lehramtsstudierenden im Praxissemester. *Zeitschrift Für Pädagogische Psychologie*, *27*(1-2), 77–86.

Gruber, H., Law, L. C., Mandl, H., & Renkl, A. (1996). Situated Learning and Transfer. In P. Reimann & H. Spada (Ed.), *Learning in Humans and Machines* (pp. 168–188). Oxford: Elsevier.

Gruber, H., & Stamouli, E. (2015). Intelligenz und Vorwissen. In E. Wild & J. Möller (Ed.), *Pädagogische Psychologie* (pp. 26–44). Berlin, Heidelberg: Springer Berlin Heidelberg.

Gump, P. V. (1967). *The Classroom Behavior Setting: Its Nature and Relation to Student Behavior*. Washington, D.C.: U.S. Department of Health, Education & Welfare, Office of Education.

Gump, P. V. (1982). School Settings and Their Keeping. In D. L. Duke (Ed.), *ASCD publication: 611-82266. Helping teachers manage classrooms* (pp. 98–114). Alexandria, Va.: Association for Supervision and Curriculum Development.

Haag, L., & Streber, D. (2012). *Klassenführung: Erfolgreich unterrichten mit Classroom Management* (1st ed). *Reihe Bildungswissen Lehramt: Vol. 23*. Weinheim, Bergstr: Beltz, J.

Hascher, T. (2006). Veränderungen im Praktikum - Veränderungen durch das Praktikum: Eine empirische Untersuchung zur Wirkung von schulpraktischen Studien in der Lehrerbildung. *Zeitschrift Für Pädagogik*, *51*(Beiheft), 130–148.

Hascher, T. (2011). Forschung zur Wirksamkeit der Lehrerbildung. In E. Terhart, H. Bennewitz, & M. Rothland (Ed.), *Handbuch der Forschung zum Lehrerberuf* (pp. 418–440). Münster: Waxmann.

Hascher, T. (2012a). Forschung zur Bedeutung von Schul- und Unterrichtspraktika in der Lehrerinnen- und Lehrerbildung. *Beiträge Zur Lehrerinnen- Und Lehrerbildung*, *30*(1), 87–98.

Hascher, T. (2012b). Lernfeld Praktikum – Evidenzbasierte Entwicklungen in der Lehrer/innenbildung. *Zeitschrift Für Bildungsforschung*, *2*(2), 109–129.

Hasselhorn, M., & Gold, A. (2013). *Pädagogische Psychologie: Erfolgreiches Lernen und Lehren* (3rd Ed.). *Standards Psychologie*. Stuttgart: Kohlhammer.

Hastie, P. A. (2006). The Classroom Ecology Paradigm. In *The Handbook of Physical Education* (pp. 214–225). London: Sage.

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

Hausknecht, J. P., Halpert, J. P., Di Paolo, N. T., & Moriarty Gerrard, M. O. (2007). Retesting in Selection: A Meta-Analysis of Practice Effects for Tests of Cognitive Ability. *Journal of Applied Psychology*, *92*, 373–385.

Havers, N. (2010). Lässt sich effiziente Klassenführung lehren?: Das Potenzial der Lehrertrainings. In J. Abel (Ed.), *Wirkt Lehrerbildung?: Antworten aus der empirischen Forschung.* Münster: Waxmann.

Hecht, M. (2009). *Selbsttätigkeit im Unterricht: Empirische Untersuchungen in Deutschland und Kanada zur Paradoxie pädagogischen Handelns* (1. Ed.). *VS Research*. Wiesbaden: VS Verlag für Sozialwissenschaften / GWV Fachverlage, Wiesbaden.

Hellermann, C., Gold, B., & Holodynski, M. (2015). Förderung von Klassenführungsfähigkeiten im Lehramtsstudium. *Zeitschrift Für Entwicklungspsychologie Und Pädagogische Psychologie*, *47*(2), 97–109.

Helmke, A. (2007). Aktive Lernzeit optimieren: Was wissen wir über Klassenführung. *Pädagogik*, *59*(5), 46–51.

Helmke, A. (2012). *Unterrichtsqualität und Lehrerprofessionalität: Diagnose, Evaluation und Verbesserung des Unterrichts* (4th ed). *Unterricht verbessern - Schule entwickeln*. Seelze-Velber: Klett/Kallmeyer.

Helmke, A., & Helmke T. (2004). Videobasierte Unterrichtsreflexion. *Thema*. (4), 48–66.

Helmke, A., & Jäger, R. S. (Ed.). (2002). *Das Projekt MARKUS: Mathematik-Gesamterhebung Rheinland-Pfalz: Kompetenzen, Unterrichtsmerkmale, Schulkontext*. Landau: Empirische Pädagogik e.V.

Helmke, A., & Renkl, A. (1992). Das Münchener Aufmerksamkeitsinventar (MAI): Ein Instrument zur systematischen Verhaltensbeobachtung der Schüleraufmerksamkeit im Unterricht. *Diagnostica*, *38*(2), 130–141.

Helsper, W. (2004). Antinomien. Widersprüche Paradoxien: Lehrerarbeit - ein unmögliches Geschäft? Eine strukturtheoretisch-rekonstruktive Perspektive auf das Lehrerhandeln. In B. Koch-Priewe, F. U.

Kolbe, & J. Wildt (Ed.), *Grundlagenforschung und mikrodidaktische Reformansätze zur Lehrerbildung* (pp. 49–98). Bad Heilbrunn: Klinkhardt.

Herzog, W. (2009). Schule und Schulklasse als soziale Systeme. In R. Becker (Ed.), *Lehrbuch. Lehrbuch der Bildungssoziologie* (p. 155). Wiesbaden: VS Verlag für Sozialwissenschaften.

Holodynski, M., Steffensky, M., Gold, B., Hellermann, C., Sunder, C., Fiebranz, A., . . . Möller, K. (2017). Lernrelevante Situationen im Unterricht beschreiben und interpretieren. In C. Gräsel & K. Trempler (Ed.), *Entwicklung von Professionalität pädagogischen Personals* (pp. 283–302). Wiesbaden: Springer Fachmedien Wiesbaden.

Holtz, P. (2014). "Es heißt ja auch Praxissemester und nicht Theoriesemester": Quantitative und qualitative Befunde zum Spannungsfeld zwischen "Theorie" und "Praxis" im Jenaer Praxissemester. In K. Kleinespel (Ed.), *Ein Praxissemester in der Lehrerbildung: Konzepte, Befunde und Entwicklungsperspektiven am Beispiel des Jenaer Modells* (pp. 97–118). Bad Heilbronn: Verlag Julius Klinkhardt.

Jackson, P. W. (1990). *Life in classrooms*. With a new introduction (2nd ed.). New York: Teachers College Press.

Jahn, G., Prenzel, M., Stürmer, K., & Seidel, T. (2011). Varianten einer computergestützten Erhebung von Lehrerkompetenzen: Untersuchungen zu Anwendungen des Tools Observer. *Unterrichtswissenschaft*, *39*(2), 136–153.

Jahn, G., Stürmer, K., Seidel, T., & Prenzel, M. (2014). Professionelle Unterrichtswahrnehmung von Lehramtsstudierenden. *Zeitschrift Für Entwicklungspsychologie Und Pädagogische Psychologie*, *46*(4), 171–180.

Jones, V. (2006). How Do Teachers Learn to Be Effective Classroom Managers? In C. M. Evertson & C. S. Weinstein (Ed.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 887–907). New York, N.J.: Routledge.

Jonkitz, E., Moosbrugger, H., & Brandt, H. (2012). Planung und Entwicklung von Tests und Fragebogen. In H. Moosbrugger & A. Kelava (Ed.), *Springer-Lehrbuch. Testtheorie und Fragebogenkonstruktion* (2nd ed.), (pp. 27–74). Berlin: Springer.

Kaiser, G., Busse, A., Hoth, J., König, J., & Blömeke, S. (2015). About the Complexities of Video-Based Assessments: Theoretical and Methodological Approaches to Overcoming Shortcomings of Research on Teachers' Competence. *International Journal of Science and Mathematics Education*, *13*(2), 369–387.

Keller, G. (2008). *Disziplinmanagement in der Schulklasse: Unterrichtsstörungen vorbeugen - Unterrichtsstörungen bewältigen* (1st ed.). *Aus dem Programm Verlag Hans Huber. Psychologie-Sachbuch*. Bern: Huber.

Kennedy, M., Soyeon, A., & Jonyoung, C. (2008). The value added by teacher education. In M. Cochran-Smith, S. Feiman-Nemser, & D. J. McIntyre (Ed.), *Handbook of research on teacher education: Enduring questions in changing contexts* (3rd ed.). New York, London: Routledge. Taylor & Francis Group.

Kersting, N. (2008). Using Video Clips of Mathematics Classroom Instruction as Item Prompts to Measure Teachers' Knowledge of Teaching Mathematics. *Educational and Psychological Measurement*, *68*(5), 845–861.

Kersting, N., Givvin, K. B., Thompson, B. J., Santagata, R., & Stigler, J. (2012). Measuring Usable Knowledge: Teachers' Analyses of Mathematics Classroom Videos Predict Teaching Quality and Student Learning. *American Educational Research Journal*, *49*(3), 568–589.

Kirschner, P., Sweller, J., & Clark, R. E. (2006). Why Minimal Guidance During Instruction does not Work: An Analysis of the Failure of Constructivist, Discorvery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist*, *41*(2), 75–86.

Klauer, K. J. (2001). Situiertes Lernen. In D. H. Rost (Ed.), *Schlüsselbegriffe. Handwörterbuch pädagogische Psychologie* (2nd ed.), (pp. 635–641). Weinheim: Beltz PVU.

Kleinknecht, M., & Schneider, J. (2013). What do teacher think and feel when analyzing videos of themselves and other teachers teaching. *Teaching and Teacher Education*, *33*(1), 13–23.

Klieme, E., & Vieluf, S. (2009) Teaching practices, teachers' beliefs and attitudes. In OECD (Ed.), *Creating Effective Teaching and Learning environments: First Results from Talis.* (pp. 87-136) *Paris:* OECD.

Klieme, E., Hartig, J., & Rauch, D. (2008). The Concept of Competence in Educational Contexts. In J. Hartig, E. Klieme, & D. Leutner (Ed.), *Assessment of Competencies in Educational Contexts: State of the Art and Future Prospects* (1st ed.). [S.l.]: Hogrefe Publishing GmbH.

KMK. (2004). *Standards für die Lehrerbildung: Bildungswissenschaften*. Retrieved from https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung.pdf

Koeppen, K., Hartig, J., Klieme, E., & Leutner, D. (2008). Current Issues in Competence Modeling and Assessment. *Zeitschrift Für Psychologie / Journal of Psychology*, *216*(2), 61–73.

König, J., Blömeke, S., Paine, L., Schmidt, W., & Hsieh, F. J. (2011). General Pedagogical Knowledge of Future Middle School Teachers: On the Complex Ecology of Teacher Education in the United States, Germany, and Taiwan. *Journal of Teacher Education*, *62*(2), 188–201.

König, J. (2013). First comes the Theory, then the Practice?: On the Acquisition of General Pedagogical Knowledge during initial teacher education. *International Journal of Science and Mathematics Education*, *11*(4), 999–1028.

König, J. (2015a). Kontextualisierte Erfassung von Lehrerkompetenzen: Einführung in den Thementeil. *Zeitschrift Für Pädagogik*, *61*(3), 305–309.

König, J. (2015b). Measuring classroom management expertise (CME) of teachers: A video-based assessment approach and statistical approach. *Cogent Education*, *2*(1).

König, J., & Blömeke, S. (2009). Pädagogisches Wissen von angehenden Lehrkräften. *Zeitschrift Für Erziehungswissenschaft*, *12*(3), 499–527.

König, J., & Blömeke, S. (2012). Future teachers' general pedagogical knowledge from a comparative perspective: does school experience matter? *ZDM*, *44*(3), 341–354.

König, J., Blömeke, S., Klein, P., Suhl, U., Busse, A., & Kaiser, G. (2014). Is teachers' general pedagogical knowledge a premise for noticing and interpreting classroom situations? A video-based assessment approach. *Teaching and Teacher Education*, *38*, 76–88.

König, J., & Kramer, C. (2015). Teacher professional knowledge and classroom management: on the relation of general pedagogical knowledge (GPK) and classroom management expertise (CME). *ZDM.* Advance online publication.

König, J., & Lebens, M. (2012). Classroom Management Expertise (CME) von Lehrkräften messen: Überlegungen zur Testung mithilfe von Videovignetten und erste empirische Befunde. *Lehrerbildung Auf Dem Prüfstand*, *5*(1), 3–28.

König, J., & Rothland, M. (2016). Klassenführungswissen als Ressource der Burnoutprävention? Zum Nutzen von pädagogisch-psychologischem Wissen im Lehrerberuf. *Unterrichtswissenschaft*, *44*(4), 425–441.

König, J., & Seifert, A. (Ed.). (2012). *Lehramtsstudierende erwerben pädagogisches Professionswissen: Ergebnisse der Längsschnittstudie LEK zur Wirksamkeit der erziehungswissenschaftlichen Lehrerausbildung*. Münster: Waxmann.

Korthagen, F. A. J. (2010). Situated learning theory and the pedagogy of teacher education: Towards an integrative view of teacher behavior and teacher learning. *Teaching and Teacher Education*, *26*(1), 98–106.

Korthagen, F. A. J., & Kessels, J. (1999). Linking Theory and Practice. *Educational Researcher*, *28*(4), 4–17.

Kounin, J. S. (1970). *Discipline and Group Management in Classrooms*. New York: Holt, Rinehart and Winston, Inc.

Kramer, C., König, J., Kaiser, G., Ligtvoet, R., & Blömeke, S. (2017). Der Einsatz von Unterrichtsvideos in der universitären Ausbildung: Zur Wirksamkeit video- und transkriptgestützter Seminare zur Klassenführung auf pädagogisches Wissen und situationsspezifische Fähigkeiten angehender Lehrkräfte. *Zeitschrift Für Erziehungswissenschaft*, *20*(S1), 137–164.

Krammer, K. (2014). Fallbasiertes Lernen mit Unterrichtsvideos in der Lehrerinnen- und Lehrerbildung. *Beitrage Zur Lehrerinnen- Und Lehrerbildung*, *32*(2), 164–175.

Krammer, K., & Reusser, K. (2004). Unterrichtsvideos als Medium der Lehrerinnen- und Lehrerbildung. *Thema*. (4), 1–22.

Krapp, A., & Weidenmann, B. (Ed.). (2006). *Anwendung Psychologie*. *Pädagogische Psychologie* (5th ed.). Weinheim: Julius Beltz & Co. KG.

Kumschick, I. R., Piwowar, V., Ophardt, D., Barth, V., Krysmanski, K., & Thiel, F. (2017). Optimierung einer videobasierten Lerngelegenheit im Problem Based Learning Format durch Cognitive Tools. Eine Interventionsstudie mit Lehramtsstudierenden. *Zeitschrift Für Erziehungswissenschaft*, *20*(S1), 93–113.

Kunina-Habenicht, O., Lohse-Bossenz, H., Kunter, M., Dicke, T., Förster, D., Gößling, J., . . . Terhart, E. (2012). Welche bildungswissenschaftlichen Inhalte sind wichtig in der Lehrerbildung?: Ergebnisse einer Delphi-Studie. *Zeitschrift Für Erziehungswissenschaft*, *15*(4), 649–682.

Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (Ed.). (2011). *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV*. Münster, New York, NY, München, Berlin: Waxmann.

Landrum, T. J., & Kauffman, J. M. (2006). Behavioral Approaches to Classroom Management. In C. M. Evertson & C. S. Weinstein (Ed.), *Handbook of classroom management: Research, practice, and contemporary issues* (Chapter 3, pp. 47–71). New York, N.J.: Routledge.

Larrivee, B. (2006). The Convergence of Reflective Practice and Efffective Classroom Management. In C. M. Evertson & C. S. Weinstein (Ed.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 983–1001). New York, N.J.: Routledge.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation / Jean Lave, Etienne Wenger*. *Learning in doing: social, cognitive, and computational perspectives*. Cambridge: Cambridge University Press.

Lefstein, A., & Snell, J. (2011). Professional Vision and the politics of teacher learning. *Teaching and Teacher Education*, *27*(3), 505–514.

Level: Lehrerbildung vernetzt entwickeln. (2016). VIGOR: Videographic Online Recorder. Retrieved from https://vigor.studiumdigitale.uni-frankfurt.de/

LimeSurvey. (2016). LimeSurvey CE. Retrieved from https://www.limesurvey.org/de

Lindmeier, A. (2013). Video-vignettenbasierte standardisierte Erhebung von Lehrerkognitionen. In U. Riegel (Ed.), *Fachdidaktische Forschungen: Vol. 4. Videobasierte Kompetenzforschung in den Fachdidaktiken* (pp. 45–61). Münster, München: Waxmann.

Lipowsky, F. (2006). Auf den Lehrer kommt es an: Empirische Evidenzen für Zusammenhänge zwischen Lehrerkompetenzen, Lehrerhandeln und dem Lernen der Schüler. In C. Allemann-Ghionda & E. Terhart (Ed.), *Zeitschrift für Pädagogik, Beiheft: Vol. 51. Kompetenzen und Kompetenzentwicklung von Lehrerinnen und Lehrern* (pp. 47–70). Beltz.

Lohmann, V., Seidel, V., & Terhart, E. (2011). Bildungswissenschaften in der universitären Lehrerbildung: Curriculare Strukturen und Verbindlichkeiten: Eine Analyse aktueller Studienordnungen an nordrhein-westfälischen Universitäten. *Lehrerbildung Auf Dem Prüfstand*, *4*(2), 271–302. Retrieved from http://www.vep-landau.de/LehrerbildungaufdemPruefstand.htm

Lüders, M. (2012). Pädagogisches Unterrichtswissen: Eine Testkritik. *Zeitschrift Für Erziehungswissenschaft*, *15*(4), 775–791.

LUQ. (2018). Lehrveranstaltungsevaluation an der Goethe Universität. Retrieved from http://www.uni-frankfurt.de/71041751/Lehrveranstaltungsevaluation

Mägdefrau, J., & Schumacher, E. (2001). Zwischen Wissen und Können: Über die Bedeutung von Erziehungswissenschaft und pädagogischer Praxis in der Lehrerbildung. *Die Deutsche Schule*, *93*(4), 411–422.

Martin, N., & Yin, Z. (2006). *Classroom Management Training, Teaching Experience and Gender: Do these Variables impact Teachers' Attitudes and Beliefs towards Classroom Management Style?* Paper presented at the Annual Conference of the Soutwest Educational Research Association, Austin, TX.

Martin, N., Yin, Z., & Baldwin, B. (1998). Construct Validation of the Attitudes & Beliefs on Classroom Control Inventory. *Journal of Classroom Interaction*, *33*(2), 6–15.

Martin, N., Yin, Z., & Mayall, H. (2007). The Attitudes & Beliefs on Classroom Control Inventory: Revised and Revisited. *Journal of Classroom Interaction*, *42*(2), 11–20.

Mason, J. (2016). Perception, interpretation and decision making: understanding gaps between competence and performance—a commentary. *ZDM.* Advance online publication.

Mayr, J. (2006). Klassenführung auf der Sekundarstufe II: Strategien und Muster erfolgreichen Lehrerhandelns. *Schweizerischer Zeitschrift Für Bildungswissenschaften*, *28*(2), 227–242.

Merrett, F., & Wheldall, K. (1993). How Do Teachers Learn to Manage Classroom Behaviour? A study of teachers' opinions about their initial training with special reference to classroom behaviour management. *Educational Studies*, *19*(1), 91–106.

Merritt, M. (1982). Distribution and Directing Attention in Primary Classrooms. In L. C. Wilkinson (Ed.), *Language, thought, and culture. Communicating in the classroom* (pp. 223–244). New York, London: Academic Press.

Mertens, S., & Gräsel, C. (2017, March). *Bildungswissenschaftliche Kompetenzentwicklung im Praxissemester: Veränderungbereiche und Einflussfakoren.* Annual Convention of the Gesellschaft für Empirische Bildungsforschung, Heidelberg.

Moosbrugger, H. (2012a). Klassische Testtheorie (KTT). In H. Moosbrugger & A. Kelava (Ed.), *Springer-Lehrbuch. Testtheorie und Fragebogenkonstruktion:* (2nd ed., Ch 5., pp. 103–117). Berlin: Springer.

Moosbrugger, H. (2012b). *Testtheorie und Fragebogenkonstruktion: Mit 66 Abbildung und 41 Tabellen* (2nd ed.). *Springer-Lehrbuch*. Berlin: Springer.

Moosbrugger, H., & Oehlschlägel, J. (2011). *Frankfurter Aufmerksamkeits-Inventar 2 (FAIR-2)*. Bern: Verlag Hans Huber.

Moosbrugger, H., Oehlschlägel, J., & Steinwascher, M. (2017). FAIR-2, Testauswerteprogramm. Göttingen: Hogrefe Verlag GmbH & Co. KG.

Moreno, R., & Valdez, A. (2007). Immediate and delayed effects of using a classroom case exemplar in teacher education: The role of presentation format. *Journal of Educational Psychology*, *99*(1), 194–206.

Müller-Fohrbrodt, G., Cloetta, B., & Dann, H.-D. (1978). *Der Praxisschock bei jungen Lehrern: Formen, Ursachen, Folgerungen e. zusammenfassende Bewertung d. theoret. u. empir. Erkenntnisse* (1st ed.). Stuttgart: Klett.

Mullis, I. V.S., Martin, M., Foy, P., & Arora, A. (2012). *TIMSS 2011: International Restuls in Mathematics*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

Mullis, I. V. S., Martin, M., Gonzalez, E. J., & Kennedy, M. M. (2003). *PIRLS 2001 International Report*. Boston: International Study Center, Lynch School of Education, Boston College.

Munk, D. D., & Repp, A. C. (1994). The Relationship between Instructional Variables and Problem Behavior: A Review. *Exceptional Children*, *60*(5), 390–401.

Neuweg, G. H. (2004). Tacit Knowing and Implicit Learning. In M. Fischer, N. Boreham, & B. Nyhan (Ed.), *Cedefop References Series. European Perspectives on Learning at Work: The Acquisition of Work Process Knowledge.* Luxemburg: Office for Official Publications for the European Communities.

Neuweg, G. H. (2011). Das Wissen der Wissensvermittler: Problemstellungen, Befunde und Perspektiven der Forschung zum Lehrerwissen. In E. Terhart, H. Bennewitz, & M. Rothland (Ed.), *Handbuch der Forschung zum Lehrerberuf* (pp. 451–477). Münster: Waxmann.

Neuweg, G. H. (2015). Kontextualisierte Kompetenzmessung: Eine Bilanz zu aktuellen Konzeptionen und forschungsmethodischen Zugängen. *Zeitschrift Für Pädagogik*, *61*(3), 377–383.

Nückles, M., & Wittwer, J. (2014). Lernen und Wissenserwerb. In T. Seidel & A. Krapp (Ed.), *Pädagogische Psychologie: Mit Online-Materialien* (6th ed., pp. 225–252). Weinheim: Beltz.

O'Neill, S., & Stephenson, J. (2012). Does classroom management coursework influence pre-service teachers' perceived preparedness or confidence? *Teaching and Teacher Education*, *28*(8), 1131–1143.

Ophardt, D., & Thiel, F. (2008). Klassenmanagement als Basisdimension der Unterrichtsqualität. In M. K. W. Schweer (Ed.), *Lehrer-Schüler-Interaktion: Inhaltsfelder, Forschungsperspektiven und methodische Zugänge* (2nd ed., pp. 259–282). Wiesbaden: VS Verlag für Sozialwissenschaften.

Ophardt, D., & Thiel, F. (2013). *Klassenmanagement: Ein Handbuch für Studium und Praxis*. *Schulpädagogik*. Stuttgart: Kohlhammer.

Oser, F. (2013). “I Know How to Do It, But I Can’t Do It”. In S. Blömeke, O. Zlatkin-Troitschanskaia, C. Kuhn, & J. Fege (Ed.), *Modeling and Measuring Competencies in Higher Education* (pp. 45–59). Rotterdam: Sense Publishers.

Oser, F., Heinzer, S., & Salzmann, P. (2010). Die Messung der Qualität von professionellen Kompetenzprofilen von Lehrpersonen mit Hilfe der Einschätzung von Filmvignetten. *Unterrichtswissenschaft*, *38*(1), 6–28.

Pajares, F. M. (1992). Teachers' Beliefs and Educational Research: Cleaning Up a Messy Construct. *Review of Educational Research*, *62*(3), 307–332.

Parsons, T. (1959). The School Class as a Social System: Some of Its Functions in American Society. *Harvard Educational Review*, *29*(4), 297–318.

Pepper, D., Hodgen, J., Lamesoo, K., Kõiv, P., & Tolboom, J. (2016). Think aloud: Using cognitive interviewing to validate the PISA assessment of student self-efficacy in mathematics. *International Journal of Research & Method in Education*, *41*(1), 3–16.

Petermann, F. (2011). Frankfurter Aufmerksamkeits--Inventar 2 (FAIR-2). *Zeitschrift Für Psychiatrie, Psychologie Und Psychotherapie*, *59*(4), 325–326.

Petko, D., Prasse, D., & Reusser, K. (2014). Online-Plattformen für die Arbeit mit Unterrichtsvideos: Eine Übersicht. *Beitrage Zur Lehrerinnen- Und Lehrerbildung*, *32*(2), 247–261.

Petko, D., & Haab, S., Reusser, K. (2003). Mediennutzung in der Lehrerinnen- und Lehrerbildung: Eine Umfrage in der deutschsprachigen Schweiz. *Beiträge Zur Lehrerinnen- Und Lehrerbildung*, *21*(1), 8–31.

Pfitzner-Eden, F., Thiel, F., & Horsley, J. (2014). An Adapted Measure of Teacher Self-Efficacy for Preservice Teachers: Exploring its Validity Across two Countries. *Zeitschrift Für Pädagogische Psychologie*, *28*(3), 83–92.

Piwowar, V., Barth, V. L., Ophardt, D., & Thiel, F. (2017). Evidence-based scripted videos on handling student misbehavior: The development and evaluation of video cases for teacher education. *Professional Development in Education*, *10*(2), 1–16.

Practical Semester Hesse. (2017). Handreichung Praxissemester. Retrieved from http://www.uni-frankfurt.de/65435609/Handreichung.pdf

Practical Semester Hesse. (2018). Evaluation, Project Description. Retrieved from https://www.uni-frankfurt.de/60154650/10_Projektbeschreibung

Putnam, R. T. (1987). Structuring and Adjusting Content for Students: A Study of Live and Simulated Tutoring of Addition. *American Educational Research Journal*, *24*(1), 13–48.

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.

Rauin, U., Herrle, M., & Engartner, T. (Ed.). (2016). *Grundlagentexte Methoden*. *Videoanalysen in der Unterrichtsforschung: Methodische Vorgehensweisen und Anwendungsbeispiele*. Weinheim: Beltz Juventa.

Rauin U. (2011). Forschung zur Lehrerbildung aus nationaler und internationaler Perspektive. In E. Terhart, H. Bennewitz, & M. Rothland (Ed.), *Handbuch der Forschung zum Lehrerberuf* (pp. 441–447). Münster: Waxmann.

Reinmann, G. & Mandl, H. (2006). Unterrichten & Lernumgebungen Gestalten. In A. Krapp & B. Weidenmann (Ed.), *Anwendung Psychologie. Pädagogische Psychologie* (5th ed.). Weinheim: Julius Beltz & Co. KG.

Renkl, A. (1996). Träges Wissen: Wenn Erlerntes nicht genutzt wird. *Psychologische Rundschau*, *47*(2), 78–92.

Rost, D. (Ed.). (2001). *Schlüsselbegriffe*. *Handwörterbuch pädagogische Psychologie* (2nd ed.). Weinheim: Beltz PVU.

Rost, J. (2004). *Lehrbuch Testtheorie, Testkonstruktion* (2nd ed.). *Aus dem Programm Huber, Psychologie Lehrbuch*. Bern: H. Huber.

Sabers, D., Cushing, K., & Berliner, D. (1991). Differences Among Teachers in a Task Characterized by Simultaneity, Multidimensionality, and Immediacy. *American Educational Research Journal*, *28*(1), 63–88.

Santagata, R., & Guarino, J. (2011). Using video to teach future teachers to learn from teaching. *ZDM*, *43*(1), 133–145.

Santagata, R., & Yeh, C. (2014). Learning to teach mathematics and to analyze teaching effectiveness: Evidence from a video- and practice-based approach. *Journal of Mathematics Teacher Education*, *17*(6), 491–514.

Santagata, R., & Yeh, C. (2016). The role of perception, interpretation, and decision making in the development of beginning teachers' competence. *ZDM*, *48*(1-2), 153–165.

Santagata, R., Zannoni, C., & Stigler, J. (2007). The role of lesson analysis in pre-service teacher education: an empirical investigation of teacher learning from a virtual video-based field experience. *Journal of Mathematics Teacher Education*, *10*(2), 123–140.

Schaarschmidt, U., Kieschke, U., & Fischer, A. W. (2017). *Lehrereignung: Voraussetzungen erkennen - Kompetenzen fördern - Bedingungen gestalten* (1st ed). *Brennpunkt Schule*. Stuttgart: Verlag W. Kohlhammer.

Fydrich, T., Moosbrugger, H., & Zielinski, W. (Ed.). (2012). *Springer-Lehrbuch*. *Psychologische Diagnostik*. Berlin, Heidelberg: Springer.

Schoenfeld, A. H. (2012). *How we Think.* XIII CIAEM, Recife. Retrieved from https://www.researchgate.net/publication/317953371_How_we_think_A_theory_of_goal-oriented_decision_making_and_its_educational_applications

Schoenfeld, A. H. (2015). How We Think: A Theory of Human Decision-Making, with a Focus on Teaching. In S. J. Cho (Ed.), *The Proceedings of the 12th International Congress on Mathematical Education* (pp. 229–243). Cham: Springer International Publishing.

Seidel, T., Blomberg, G. A., & Renkl, A. (2013). Instructional strategies for using video in teacher education. *Teaching and Teacher Education*, *34*, 56–65.

Seidel, T., Blomberg, G. A., & Stürmer, K. (2010). "Observer": Validierung eines videobasierten Instruments zur Erfassung der professionellen Wahrnehmung von Unterricht. Projekt OBSERVE. *Zeitschrift Für Pädagogik*, *56*(Beiheft), 296–306.

Seidel, T., & Krapp, A. (Ed.). (2014). *Pädagogische Psychologie: Mit Online-Materialien* (6th ed.). Weinheim: Beltz.

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.

Seidel, T., & Stürmer, K. (2014). Modeling and Measuring the Structure of Professional Vision in Preservice Teachers. *American Educational Research Journal*, *51*(4), 739–771.

Seidel, T., Stürmer, K., Blomberg, G. A., 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.

Seidel, T., & Thiel, F. (2017). Standards und Trends der videobasierten Lehr-Lernforschung. *Zeitschrift Für Erziehungswissenschaft*, *20*(S1), 1–21.

Seidel T. (2009). Klassenführung. In E. Wild (Ed.), *Springer-Lehrbuch. Pädagogische Psychologie* (1st ed., Chapter 6, pp. 135–148). Berlin: Springer.

Seiz, J., Voss, T., & Kunter, M. (2015). When Knowing is Not Enough - the Relevance of Teachers' Cognitive and Emotional Resources for Classroom Management. *Frontline Learning Research, 3(1)*, 55-77.

Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (Ed.). *Standards für die Lehrerbildung: Bildungswissenschaften: Beschluss der Kulturministerkonferenz vom 16.12.2004*. Bonn.

Shavelson, R. (2010). On the measurement of competency. *Empirical Research in Vocational Education and Training*, *2*(1), 41–63.

Sherin, M. G. (2001). Developing a Professional Vision of Classroom Events. In T. L. Wood, B. S. Nelson, & J. Warfield (Ed.), *Studies in mathematical thinking and learning. Beyond classical pedagogy: Teaching elementary school mathematics* (pp. 75–93). Mahwah, N.J: L. Erlbaum Associates.

Sherin, M. G. (2007a). The Development of Teachers' Professional Vision in Video Clubs. In R. Goldman, R. D. Pea, B. Barron, & S. Derry (Ed.), *Video Research in the Learning Sciences* (pp. 383–395). Mahwah, N.J.: Lawrence Erlbaum Associates.

Sherin, M. G. (2007b). New Perspectives on the Role of Video in Teacher Education. In J. E. Brophy (Ed.), *Advances in research on teaching: Vol. 10. Using video in teacher education* (1st ed.), (pp. 1–27). Amsterdam, London: JAI.

Sherin, M. G., & Han, S. Y. (2004). Teacher learning in the context of a video club. *Teaching and Teacher Education*, *20*(2), 163–183.

Sherin, M. G., Jacobs, V. R., & Philipp, R. A. (Ed.). (2011a). *Studies in mathematical thinking and learning*. *Mathematics Teacher Noticing: Seeing through Teachers' Eyes*. London: Routledge.

Sherin, M. G., Jacobs, V. R., & Philipp, R. A. (2011b). Situating the study of Teacher Noticing. In M. G. Sherin, V. R. Jacobs, & R. A. Philipp (Ed.), *Studies in mathematical thinking and learning. Mathematics Teacher Noticing: Seeing through Teachers' Eyes* (pp. 3–13). London: Routledge.

Sherin, M. G., & Van Es, E. A. (2005). Using Video to Support Teachers' Ability to Notice Classroom Interactions. *Journal of Technology and Teacher Education*, *13*(3), 475–491.

Sherin, M. G., & Van Es, E. A. (2009). Effects of Video Club Participation on Teachers' Professional Vision. *Journal of Teacher Education*, *60*(1), 20–37.

Shulman, L. S. (1986). Those Who Understand: Knowledge Growth in Teaching. *Educational Researcher*, *15*(2), 4–14.

Shulman, L. S. (1987). Knowledge and teaching: Foundations for a new reform. *Harvard Educational Review*, *57*(1), 1–23.

Shulman, L. S. (1992). Toward a Pedagogy of Cases. In L. S. Shulman (Ed.), *Case Methods in Teacher Education* (pp. 1–30). New York, NY: Teachers College Press.

Slavin, R. E. (1994). Quality, appropriateness, incentive, and time: A model of instructional effectiveness. *International Journal of Educational Research*, *21*(2), 141–157.

Stahnke, R., Schueler, S., & Roesken-Winter, B. (2016). Teachers' perception, interpretation, and decision-making: A systematic review of empirical mathematics education research. *ZDM*, *48*(1-2), 1–27.

Star, J. R., Lynch, K., & Perova, N. (2011). Using Video to Improve Preservice Mathematics Teachers' Abilities to attend to Classroom features: A Replication study. In M. G. Sherin, V. R. Jacobs, & R. A. Philipp (Ed.), *Studies in mathematical thinking and learning. Mathematics Teacher Noticing: Seeing through Teachers' Eyes* (pp. 117–133). London: Routledge.

Star, J. R., & Strickland, S. K. (2008). Learning to observe: using video to improve preservice mathematics teachers' ability to notice. *Journal of Mathematics Teacher Education*, *11*(2), 107–125.

Steffensky, M., Gold, B., Holdynski, M., & Möller, K. (2015). Professional Vision of Classroom Management and Learning Support in Science Classrooms—Does Professional Vision Differ Across General and Content-Specific Classroom Interactions? *International Journal of Science and Mathematics Education*, *13*(2), 351–368.

Steffensky, M., & Kleinknecht, M. (2016). Wirkungen videobasierter Lernumgebungen auf die professionelle Kompetenz und das Handeln (angehender) Lehrpersonen: Ein Überblick zu Ergebnissen aus aktuellen (quasi-)experimentellen Studien. *Unterrichtswissenschaft*, *46*(4), 305–322.

Stough, L. M. (2006). The Place of Classroom Management and Standards in Teacher Education. In C. M. Evertson & C. S. Weinstein (Ed.), *Handbook of classroom management: Research, practice, and contemporary issues* (Chapter 34, pp. 909–923). New York, N.J.: Routledge.

Stürmer, K., & Seidel, T. (2015). Assessing Professional Vision in Teacher Candidates. *Zeitschrift Für Psychologie*, *223*(1), 54–63.

Stürmer, K., & Seidel, T. (2017). A Standardized Approach for Measuring Teachers' Professional Vision: The Observer Research Tool. In E. O. Schack, M. H. Fisher, & J. A. Wilhelm (Ed.), *Teacher Noticing: Bridging and Broadening Perspectives, Contexts, and Frameworks* (pp. 359–380). Cham: Springer International Publishing.

Stürmer, K., Seidel, T., & Schäfer, S. (2013). Changes in professional vision in the context of practice. *Gruppendynamik Und Organisationsberatung*, *44*(3), 339–355.

Sun, J., & Van Es, E. A. (2015). An Exploratory Study of the Influence That Analyzing Teaching Has on Preservice Teachers' Classroom Practice. *Journal of Teacher Education*, *66*(3), 201–214.

Sweller, J. (1988). Cognitive Load During Problem Solving: Effects on Learning. *Cognitive Science*, *12*(2), 257–285.

Sweller, J. (2002). Visualisation and instructional design. In *Proceedings of the International Workshop on Dynamic Visualizations and Learning* (pp. 1501-1510).

Sweller, J., Van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review*, *10*(3), 251–296.

Syring, M., Kleinknecht, M., Bohl, T., Kuntze, S., Rehm, M., & Schneider, J. (2015). How Problem-Based or Direct Instructional Case-Base Learning Environments influence Secondary School Preservice Teachers' Cognitive Load, Motivation and Emotions: A Quasi-Experimental Intervention Study in Teacher Education. *Journal of Education and Human Development*, *4*(4), 115–129.

Tenorth, H.E. (2006). Professionalität im Lehrerberuf: Ratlosigkeit der Theorie, gelingende Praxis. *Zeitschrift Für Erziehungswissenschaft*, *9*(4), 580–597.

Tenorth, H.E. (2010). *Geschichte der Erziehung: Einführung in die Grundzüge ihrer neuzeitlichen Entwicklung* (5th Edition). *Grundlagentexte Pädagogik*. Weinheim, München: Juventa.

Terhart, E., Bennewitz, H., & Rothland, M. (Ed.). (2011). *Handbuch der Forschung zum Lehrerberuf*. Münster: Waxmann.

Terhart, E., Schulze-Stocker, F., Holzberger, D., & Kunina-Habenicht, O. (2013). Bildungswissenschaften in der universitären Lehrerbildung: Unterscheiden sich Absolventinnen und Absolventen mit Erstem Staatsexamen von Absolventinnen und Absolventen mit einem Master of Education-Abschluss? In A. Gehrmann (Ed.), *Formation und Transformation der Lehrerbildung: Entwicklungstrends und Forschungsbefunde* (pp. 224–238). Bad Heilbrunn: Klinkhardt.

Thiel, F., Ophardt, D., & Piwowar, V. (2013). Kompetenzen des Klassenmanagements (KODEK): Entwicklung und Evaluation eines Fortbildungsprogramms für Lehrkräfte zum Klassenmanagement. Abschlussbericht des Projekts. Retrieved from http://www.ewi-psy.fu-berlin.de/einrichtungen/arbeitsbereiche/schulentwicklungsforschung/downloads/Abschlussbericht-KODEK.pdf

Tillmann, K.J. (2016). Empirische Bildungsforschung in der Kritik – ein Überblick über Themen und Kontroversen. *Zeitschrift Für Erziehungswissenschaft*, *19*(S1), 5–22.

Van Es, E. A., & Sherin, M. G. (2002). Learning to Notice: Scaffolding New Teachers' Interpretations of Classroom Interactions. *Journal of Technology and Teacher Education*, *10*(4), 571–596.

Van Es, E. A., & Sherin, M. G. (2008). Mathematics teachers' "learning to notice" in the context of a video club. *Teaching and Teacher Education*, *24*(2), 244–276.

Van Merriënboer, J. J. G., & Kirschner, P. (2009). *Ten steps to complex learning: A systematic approach to four-component instructional design*. New York, NY: Routledge, Taylor & Francis Group.

Van Merriënboer, J. J. G., Krischner, P. A., & Kester, L. (2003). Taking the Load Off a Learner's Mind: Instructional Design for Complex Learning. *Educational Psychologist*, *38*(1), 5–13.

Veenman, S. (1984). Perceived Problems of Beginning Teachers. *Review of Educational Research*, *54*(2), 143–178.

Voss, T., Kleickmann, T., Kunter, M., & Hachfeld, A. (2011). Überzeugungen von Mathematiklehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Ed.), *Professionelle*

Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV (pp. 235–257). Münster, New York, NY, München, Berlin: Waxmann.

Voss, T., Kunina-Habenicht, O., Hoehne, V., & Kunter, M. (2015). Stichwort Pädagogisches Wissen von Lehrkräften: Empirische Zugänge und Befunde. *Zeitschrift Für Erziehungswissenschaft.* Advance online publication.

Voss, T., Kunter, M., & Baumert, J. (2011). Assessing teacher candidates' general pedagogical/psychological knowledge: Test construction and validation. *Journal of Educational Psychology*, *103*(4), 952–969.

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1993). Toward a Knowledge Base for School Learning. *Review of Educational Research*, *63*(3), 249–294.

Weinert, F. E. (2001). A Concept of Competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Ed.), *Defining and selecting key competencies* (pp. 45–65). Kirkland, WA: Hogrefe & Huber.

Wirtz, M. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen*. Göttingen: Hogrefe.

Wolff, C., Van den Bogert, N., Jarodzka, H., & Boshuizen, H. P. A. (2015). Keeping an Eye on Learning: Differences Between Expert and Novice Teachers' Representations of Classroom Management Events. *Journal of Teacher Education*, *66*(1), 68–85.

Wolff, C., Jarodzka, H., & Boshuizen, H. P. A. (2017). See and tell: Differences between expert and novice teachers' interpretations of problematic classroom management events. *Teaching and Teacher Education*, *66*, 295–308.

Wolff, C., Jarodzka, H., Van den Bogert, N., & Boshuizen, H. P. A. (2016). Teacher vision: Expert and novice teachers' perception of problematic classroom management scenes. *Instructional Science*, *44*(3), 243–265.

Woolfolk Hoy, A. (2005). *Educational psychology* (9th ed., Active learning ed.). Boston: Pearson/A and B.

Woolfolk Hoy, A., & Weinstein, C. S. (2006). Student and Teacher Perspectives on Classroom Management. In C. M. Evertson & C. S. Weinstein (Ed.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 181–219). New York, N.J.: Routledge.

Zeichner, K. (2009). Rethinking the Connections Between Campus Courses and Field Experiences in College- and University-Based Teacher Education. *Journal of Teacher Education*, *61*(1-2), 89–99.

Zymek, B. (2011). Grundzüge der Schulentwicklung in Deutschland im 19. und 20. Jahrhundert. In S. Hellekamps, W. Plöger, & W. Wittenbruch (Ed.), *Handbuch der Erziehungswissenschaft. Schule: Handbuch der Erziehungswissenschaft 3* (1st ed.), (pp. 65–84). Stuttgart: UTB.

APPENDICES

Appendix A: CMC (Preliminary Version), Items, Criteria, and Expert-Ratings

Table 21, part I

CMC (Preliminary Version) Items, and Criteria.

Item				Criterion	
ID	Item	SSS	CM	ID	Criterion
1.1	Nennen Sie die von Ihnen beobachteten Handlungsmaßnahmen (in Stichworten), mit denen die Lehrerin versucht die Aufmerksamkeit der SchülerInnen auf sich zu lenken.	AP	LP	1.1K1	Stellt sich in die Mitte des Raumen
				1.1K2	Gruppenansage: 'Boys and girls' & 'I'm waiting'
				1.1K3	'Shh-laute'
				1.1K4	Gestikuliert
				1.1K5	Klatschen
				1.1K6	Stille
				1.1K7	Direktes Ansprechen einer S.
				1.1K8	Satzunterbrechen
				1.1K9	Blickkontakt störende SuS
1.2	Welche Handlungsmaßnahme der Lehrerin war im Videoabschnitt besonders erfolgreich für die Herstellung der Aufmerksamkeit der ganzen Klasse	HP	LP	1.2K1	Nennen des Namens einer Schülerin
1.3	Wie viele SuS sitzen während der Arbeitsphase sichtbar mit dem Rücken zur Lehrerin? A: keine/n B: 1 oder 2 C: 3 oder 4 D: 5 oder mehr6	AP	CA	1.2K1	D: 5 oder mehr
1.4	Wie viele SuS hat die Lehrerin insgesamt persönlich mit Namen angesprochen A: keine/n B: 1 C: 2 D: 3	AP	LP	1.4K1	B: 1
1.5	Warum ist es der Lehrerin so wichtig, dass sie die Aufmerksamkeit aller SuS bekommt?	JA	LP	1.3K1	Keine Wiederholung
				1.3K2	Vermeidung Verständnisfragen
				1.3K3	Es entstehen ablenkende Geräusche
				1.3K4	So verliert sie nicht ihre Autorität

Note. SSS = situation-specific skills; CM = dimension of classroom management; AP = accuracy of perception; HP, holistic perception; JA = justification of action; CA = classroom arrangement; RR = rules and routines; LP = Low Profile;)

Table 21, part II

Item				Criterion	
ID	Item	SSS	CM	ID	Criterion
1.6	Welches Problem versucht die Lehrerin zu vermeiden, indem sie sagt, nur ein/e SciiherIn pro Gruppe solle zur Tafel kommen?	JA	RO	1.6K1 1.6K2	Sie versucht Zeit zu sparen Sie versucht Störungen zu vermeiden
1.7	Die Lehrerin muss sich zweimal Mühe geben, die Aufmerksamkeit der SuS zu gewinnen. Welche der folgenden Aussagen trifft am besten auf diese zwei Phasenübergänge zu? A: Beim ersten Phasenübergang sind die SuS eher mit unterrichtsbezogenen Aktivitäten beschäftigt, beim zweiten Phasenübergang eher mit unterrichtsfernen Aktivitäten B: Beim ersten Phasenübergang sind die SuS eher mit unterrichtsfernen Aktivitäten beschäftigt, beim zweiten Phasenübergang eher mit unterrichtsbezogenen Aktivitäten C: Bei beiden Phasenübergänge sind die SuS eher mit Unterrichtsbezogenen Aktivitäten beschäftigt. D: Bei beiden Phasenübergänge sind die SuS eher mit Unterrichtsfernen Aktivitäten beschäftigt.	HP	LP	1.7K1	A
1.8	Welche Aufforderung der Lehrerin führt dazu, dass die SuS erneut unruhig werden?	HP	LP	1.8K1	Das Aufhängen der Plakate
2.1	Warum unterbricht die Lehrerin ihre Einführung und geht zur anderen Seite des Klassenraums? Wählen Sie die Aussage, die am besten zutrifft? A: Sie hilft einem Schüler, sich vom Nachbarn zu entfernen B: Sie entfernt einen Gegenstand, den sie als potentielle Störung einschätzt C: Sie tadelt einen Schüler, der nicht aufpasst. D: Sie öffnet das Fenster.	JA	RO	2.1K1	A
2.2	Welchen Grund nennt die Lehrerin dafür, dass ein Schüler von seinem Mitschüler wegrücken solle?	HP	RO	2.2K1	Das Kind drängelt ihn

Note. SSS = situation-specific skills; CM = dimension of classroom management; AP = accuracy of perception; HP, holistic perception; JA = justification of action; CA = classroom arrangement; RR = rules and routines; LP = Low Profile

Table 21, part III

Item				Criterion	
ID	Item	SSS	CM	ID	Criterion
2.3	Welchen Grund könnte es haben, das die Lehrerin das Daum schon vor Unterrichtsbeginn an die Tafel geschrieben hat?	JA	RR	2.3K1 2.3K2 2.3K3 2.3K4	Das ist Routine/ macht sie immer so Sie möchte der Klasse nicht den Rücken zudrehen Sie möchte die Aufmerksamkeit der SuS bereits damit lenken Sie möchte Zeit sparen
2.4	Welche der folgenden Aussagen trifft am besten auf die Sitzordnung zu? A: Alle SuS sitzen in Zweiergruppen B: Alle SuS sitzen in Vierergruppen C: Alle SuS sitzen in Vierer-, Fünfer, oder Sechsergruppen. D: Die Sus sitzen sowohl in Zweier- als auch in Vierer, Fünfer, oder Sechsergruppen.	AP	RO	2.4K1	D
2.5	Nennen Sie die von Ihnen beobachteten Handlungsmaßnahmen, mit denen die Lehrerin das unaufgeforderte Rufen des Datums einiger SuS unterbindet?	AP	LP	2.5K1 2.5K2 2.5K3	Die Lehrerin ignoriert sie Die Lehrerin legt ihren Finger auf den Mund Die Lehrerin sagt ‚it's not your turn'
2.6	Welchen im Video gesehenen Probleme will die Lehrerin entgegenwirken, in dem sie versucht, das unaufgeforderte Rufen des Datums zu unterbinden?	JA	LP	2.6K1 2.6K2 2.6K3 2.6K4	Chaos Zum Schutz von schwächeren Kindern Verwirrung Regelverletzung
2.7	Die Lehrerin fragt einen S insgesamt zweimal nach dem Datum. Aus welchem Grund nimmt sie ihn das zweite Mal dran?	JA	LP	2.7K1	Er hat nicht aufgepasst
2.8	Woran kann man erkennen, dass es den SuS erlaubt ist während der Stunde zu trinken? A: Es stehen Trinkbecher auf den Tischen. B: Es stehen Wasserflaschen auf den Tischen. C: Sie antwortet dies, wenn ein S. danach fragt D: Einige SuS trinken im Blickfeld der Lehrerin	AP	RR	2.8K1	A
3.1	Die Lehrerin setzt sich im Türbereich mit einigen Ereignissen auseinander, bevor sie mit dem Grußaustausch beginnt. Markieren Sie alle Ereignisse, die auf den gesehenen Abschnitt zutreffen * Ein S, der nicht zur Klasse gehört, steht in der Türöffnung * Ein S hat noch Kaugummi im Mund. * Ein S hat ein persönliches Anliegen, das er mit der Lehrerin besprechen möchte. * Die Lehrerin geht kurz auf den Flur * Ein S hat noch eine Kappe auf dem Kopf	AP	LP	3.1K1 3.1K2 3.1K3 3.1K4 3.1K5	Trifft zu Trifft zu Trifft zu Trifft nicht zu Trifft nicht zu

Note. SSS = situation-specific skills; CM = dimension of classroom management; AP = accuracy of perception; HP, holistic perception; JA = justification of action; CA = classroom arrangement; RR = rules and routines; LP = Low Profile

Table 21, part IV

Item				Criterion	
ID	Item	SSS	CM	ID	Criterion
3.2	Aus welchem Grund gibt die Lehrerin sich so viel Mühe, die Ereignisse im Türbereich schnell zu beenden?	JA	RR	3.2K1 3.2K2	Sie möchte schnell mit der Stunde anfangen Sie möchte potentielle Störungen unterbinden
3.3	Aus welchen Aussagen der Lehrerin kann man schließen, dass die SuS feste Sitzplätze haben?	HP	RO	3.3K1 3.3K2	Sie sagt an: ‚geht auf euren Plätzen. Sie fragt an die Klasse: ‚sitzt ihr wieder wo anders?‘
3.4	Wie viele SuS werden von der Lehrerin persönlich aufgefordert aufzustehen? A: keine/r B: 1 C: 2 D: 3	AP	LP	3.4K1	C
3.5	Welchen Sinn könnte es haben, dass die Lehrerin alle SuS auffordert aufzustehen, bevor sie mit dem Grußaustausch anfängt?	JA	RR	3.5K1 3.5K2 3.5K3 3.5K4 3.5K5	Sie hat schnell die Übersicht, wer mitmacht Sie aktiviert die SuS Sie fordert die Homogenität und das Gruppengefühl Die fordert Disziplin und Höflichkeit Es zeigt eine klare Markierung des Stundenanfangs
3.6	Wählen Sie die Aussage, die am besten auf das SuSverhalten am Anfang und am Ende des Grußaustausches zutrifft. A: Am Anfang und am Ende stehen alle SuS B: Am Anfang und am Ende sitzen alle SuS C: Am Anfang sitzen alle SuS und am Ende stehen alle SuS D: Am Anfang stehen alle SuS und am Ende sitzen alle SuS	AP	RR	3.6K1	D
3.7	Nennen Sie einen Vorteil der von Ihnen beobachteten Tischordnung hinsichtlich der Prävention von Unterrichtsstörungen durch die SchülerInnen.	HP	RO	3.7K1 3.7K2 3.7K3 3.7K4	Die SuS wenden sich der Lehrerin zu Die SuS schauen einander auf den Rücken Die SuS haben einen größeren Abstand zu einander Die SuS haben maximal eine/n Sitznachbar
3.8	Benennen Sie in Stichworten die Wirkungen des Grußaustausches auf das Verhalten der SchülerInnen.	HP	RR	3.8K1 3.8K2 3.8K3 3.8K4 3.8K5	Die SuS sitzen alle an ihrem Platz Die SuS werden leise Die SuS schauen alle auf die Lehrerin Die SuS haben die Aufmerksamkeit auf den Unterricht Die SuS werden aktiviert
4.1	Benennen Sie in Stichworten das Verhalten, mit dem der Lehrer eine neue Unterrichtsphase einleitet.	AP	RR	4.1K1 4.1K2 4.1K3 4.1K4	Der Lehrer sagt: ‚So…‘ Der Lehrer positioniert sich zentral Der Lehrer hebt seine Hand Der Lehrer sagt ,,3, 2, 1, stopp‘

Note. SSS = situation-specific skills; CM = dimension of classroom management; AP = accuracy of perception; HP, holistic perception; JA = justification of action; CA = classroom arrangement; RR = rules and routines; LP = Low Profile

Table 21, part V

Item				Criterion	
ID	Item	SSS	CM	ID	Criterion
4.2	Nennen Sie in Stichworten die Handlungsmaßnahmen, mit denen der Lehrer die SchülerInnen zum Flüstern bewegt.	AP	LP	4.2K1	Er verweist auf Routine: ‚Wie wir schon oft gemacht haben'
				4.2K2	Er fängt selber an zu flüstern
				4.2K3	Er sagt: ‚whisper'
				4.2K4	Er hebt beide Hände
				4.2K5	Er sagt ‚shhh'
4.3	Welchen Vorteil hat die gewählte Arbeitsform in Bezug auf potentielle Unterrichtsstörungen durch die SchülerInnen.	HP	RR	4.3K1	Die SuS schauen ins Buch/ jeder für sich
				4.3K2	Sie SuS reden leise
				4.3K3	Die SuS reden mit dem Tonband und nicht mit einander
				4.3K4	Eventuelle Störer fallen schneller auf
4.4	Welche der folgenden Aussagen kennzeichnet die Sitzordnung am besten A: In den vordersten Reihen sitzen nur Jungs B: In den vordersten Reihen sitzen nur Mädchen C: In den vordersten Reihen sitzen Jungs und Mädchen gemischt	AP	RO	4.4K1	A
4.5	Warum geht der Lehrer während der Übung zu einer Schülerin? Wählen Sie, was am besten zutrifft A: Er will nicht, dass sie sich dauernd zur Sitznachbarin umdreht B: Er will nicht, dass sie ihr Buch in die Luft hält C: Er will nicht, dass sie andere Materialien vor sich hat. D: Er will nicht, dass sie dauernd mit der Sitznachbarin quatscht	JA	LP	4.5K1	C
4.6	Aus welchem Grund entscheidet der Lehrer erst zum Tisch der Schülerin zu gehen, bevor er sie tadelt?	JA	RO	4.6K1	Er möchte die anderen SuS nicht stören
				4.6K2	Er möchte die Schülerin nicht bloßstellen
				46K3	Es ist ihm unklar, was los ist
				4.6K4	Die körperliche Nähe erzeugt Dominanz

Note. SSS = situation-specific skills; CM = dimension of classroom management; AP = accuracy of perception; HP, holistic perception; JA = justification of action; CA = classroom arrangement; RR = rules and routines; LP = Low Profile

Appendix B: CMC (Final Version), Items, Criteria, and Expert-Ratings

Table 22, part I

***CMC (Final Version), Items, Criteria, Mean, and Standard Deviations of Ratings of the Second Round of Expert Ratings and the Value of the Third Round of Expert-Ratings*.**

Item				Criterion		Expert-Rating[a]		
ID	Item	SSS	CM	ID	Criterion	M	SD	*3*
1.1	Nennen Sie die von Ihnen beobachteten Handlungsmaßnahmen (in Stichworten). mit denen die Lehrerin versucht die Aufmerksamkeit der SchülerInnen auf sich zu lenken.	AP	LP	1.1K1	Änderung der Raumpositionierung	X	X	2
				1.1K2	Explizite Handlungsanweisungen	4.2	1.3	1
				1.1K3	'Shh-laute'.	5.0	0.0	1
				1.1K4	Gestikuliert	4.8	0.4	1
				1.1K5	Klatschen.	4.0	1.6	2
				1.1K6	Stille	4.4	1.0	2
				1.1K7	Direktes Ansprechen	3.1	1.5	1
				1.1K8	Erhebung der Stimme	3.5	1.0	2
				1.1K9	Blickkontakt störende SuS	3.9	1.0	2
				1.1K10	Allgemeine Gruppenansagen	4.8	0.7	1
1.2	*Wie viele SuS sitzen während der Arbeitsphase sichtbar mit dem Rücken zur Lehrerin?* *A: keine/n* *B: 1 oder 2* *C: 3 oder 4* *D: 5 oder 6* *E: 7 oder 8*	*AP*	*CA*	*1.2K1*	*D: 5 oder 6*	*X*	*X*	*X*
1.3	Welche Probleme versucht die Lehrerin zu vermeiden. indem sie sagt. nur ein/e SchülerIn pro Gruppe solle zur Tafel kommen?	JA	CA	1.3K1	Zeitsparen	3.2	1.1	2
				1.3K2	Unruhe	4.0	1.1	1
				1.3K3	Platzmangel	4.2	0.8	1
				1.3K4	Verlust des Überblickes	3.1	1.2	1

Note. SSS = situation-specific skills; CM = dimension of classroom management; AP = accuracy of perception; HP, holistic perception; JA = justification of action; CA = classroom arrangement; RR = rules and routines; LP = Low Profile; 3 = Value of third round of expert-ratings.

[a] = Results of 2nd and 3rd round of expert-ratings, X indicates that experts did not rate this criterion.

Italics indicate that an item or Criterion was deleting during the elimination process (see chapter 7.2.4)

Table 22, part II

Item				Criterion		Expert-Rating[a]		
ID	Item	SSS	CM	ID	Criterion	M	SD	*3*
1.4	*Die Lehrerin muss sich zweimal Mühe geben. die Aufmerksamkeit der SchülerInnen zu gewinnen. Welche der folgenden Aussagen trifft am besten auf diese zwei Phasenübergänge zu? A: Beim ersten Phasenübergang sind die SchülerInnen eher mit unterrichtsbezogenen Aktivitäten beschäftigt. beim zweiten Phasenübergang eher mit unterrichtsfernen Aktivitäten. B: Beim ersten Phasenübergang sind die SchülerInnen eher mit unterrichtsfernen Aktivitäten beschäftigt. beim zweiten Phasenübergang eher mit unterrichtsbezogenen Aktivitäten. C: Bei beiden Phasenübergänge sind die SchülerInnen eher mit Unterrichtsbezogenen Übungen beschäftigt D: Bei beiden Phasenübergänge sind die SchülerInnen eher mit unterrichtsfernen Aktivitäten beschäftigt.*	*HP*	*LP*	*1.4K1*	*A: Beim ersten Phasenübergang sind die SchülerInnen eher mit unterrichtsbezogenen Aktivitäten beschäftigt. beim zweiten Phasenübergang eher mit unterrichtsfernen Aktivitäten.*	*X*	*X*	*X*
1.5	*Welche Handlungsmaßnahme der Lehrerin war am erfolgreichsten für die Herstellung der Aufmerksamkeit der ganzen Klasse beim zweiten Phasenübergang?*	*HP*	*LP*	*1.5K1*	*Direkte Ansprache*	*3.8*	*1.0*	*1*
				1.5K2	*Stille*	*2.8*	*1.3*	*2*
1.6	Wie hätte die Lehrerin den Übergang zur Ergebnissicherung an der Tafel so gestalten können. dass sie insgesamt nur einmal hätte warten müssen?	HP	LP	1.6K1	Sie hätte die SuS früher instruieren können	3.2	1.4	1
				1.6K2	*SuS hätten die Ergebnisse am Sitzplatz präsentieren können*	*2.8*	*1.4*	*1*
				1.6K3	L. hätte Plakate Abholen können	2.6	1.4	1
				1.6K4	SuS-Präsentation direkt beim Aufhängen.	3.4	1.2	1
				1.6K5	Komplette Darstellung Handlungsablauf.	4.5	1.1	2
1.7	*Wie viele SchülerInnen hat die Lehrerin insgesamt persönlich mit Namen angesprochen? A: keine/n B: 1 C: 2 D: 3 E: 4*	*AP*	*LP*	*1.7K1*	*B:1*	*X*	*X*	X

Note. SSS = situation-specific skills; CM = dimension of classroom management; AP = accuracy of perception; HP, holistic perception; JA = justification of action; CA = classroom arrangement; RR = rules and routines; LP = Low Profile; 3 = Value of third round of expert-ratings.

[a] = Results of 2nd and 3rd round of expert-ratings, X indicates that experts did not rate this criterion.

Italics indicate that an item or Criterion was deleting during the elimination process (see chapter 7.2.4)

Table 22, part III

Item				Criterion		Expert-Rating[a]		
ID	Item	SSS	CM	ID	Criterion	M	SD	*3*
1.8	Die Lehrerin gibt an. dass noch wenig Zeit übrig ist. Aus welchen Gründen hält sich die Lehrerin trotzdem an der Regel. erst mit der Instruktion anzufangen. wenn sie die Aufmerksamkeit aller SchülerInnen hat?	JA	RR	1.8K1	Zeitsparen.	3.9	1.0	2
				1.8K2	*SuS Gegenseitige Ablenkun*	*4.1*	*0.8*	*1*
				1.8K3	*Untergrabung Authorität.*	*3.2*	*1.2*	*1*
				1.8K4	*Stimme schonen.*	*3.7*	*0.7*	*1*
				1.8K5	Instruktion muss wahrnehmbar sein	4.7	0.5	2
				1.8K6	Vermeiden Inkonsequenz.	4.1	1.0	1
2.1	*Welche potentiellen Störungsquellen behebt die Lehrerin. bevor sie mit der ersten Übung anfängt?*	*AP*	*CA*	*2.1K1*	*L. entfernt eine Wasserflasche*	*4.6*	*1.1*	*2*
				2.1K2	*L. fordert einen Schüler auf vom Mitschüler wegzurücken).*	*4.9*	*0.3*	*2*
				2.1K3	*L. beendet SuS-Gespräche*	*3.7*	*1.3*	*1*
2.2	In welchen Hinsichten ist es der Lehrerin für den Start der Stunde hilfreich. dass das Datum schon an der Tafel steht?	JA	LP	2.2K1	L. muss sich nicht umdrehen	4.1	1.1	1
				2.2K2	*L. richtet vor Unterricht die Aufmerksamkeit*	*3.8*	*1.1*	*1*
				2.2K3	L. muss nicht mehr suchen / Fragen	3.5	1.2	1
				2.2K4	Zeit sparen	3.6	1.3	2
				2.2K5	Routine zum Start des Unterrichts	4.3	0.9	1
2.3	Geben Sie von den unterstehenden Aussagen an. ob sie auf die Sitzordnung zutreffen	AP	CA	2.3A	Alle Sitzplätze sind belegt	X	X	X
				2.3B	Es sitzen maximal vier SchülerInnen pro Tischgruppe	X	X	X
				2.3C	SchülerInnen sitzen sowohl in zweier- als auch in größeren Gruppen	X	X	X
				2.3D	Tische sind alle so platziert. dass die Lehrerin von Ihrer Position bei der Tafel keine SchülerInnen auf den Rücken schauen muss	X	X	X
				2.3E	*Jungs sitzen immer bei/ neben Jungs und Mädchen immer bei/ neben Mädchen*	*X*	*X*	*X*
2.4	Mit welchen Handlungsmaßnahmen versucht die Lehrerin das unaufgeforderte Rufen des Datums einiger SchülerInnen zu unterbinden?	AP	LP	*2.4K1*	*Ignorierung*	*3.5*	*1.5*	*2*
				2.4K2	Gestikulieren	4.9	0.3	1
				2.4K3	Persönliche Ansprache	4.9	0.3	1
				2.4K4	L. wählt S. aus der/die das Datum nennen darf	4.8	0.4	1
				2.4K5	Shh-Laute	4.7	0.8	1

Note. SSS = situation-specific skills; CM = dimension of classroom management; AP = accuracy of perception; HP, holistic perception; JA = justification of action; CA = classroom arrangement; RR = rules and routines; LP = Low Profile; 3 = Value of third round of expert-ratings.

[a] = Results of 2nd and 3rd round of expert-ratings, X indicates that experts did not rate this criterion.

Italics indicate that an item or Criterion was deleting during the elimination process (see chapter 7.2.4)

Table 22, part IV

Item				Criterion		Expert-Rating[a]		
ID	Item	SSS	CM	ID	Criterion	M	SD	3
2.5	Wie viele SchülerInnen haben sich während der Übung per Handhaben gemeldet? A: keine/n B: 1 C: 2 D: 3 E: 4	AP	RR	2.5K1	C. 2	X	X	X
2.6	Die Lehrerin fragt einen Schüler insgesamt zweimal nach dem Datum. Was will die Lehrerin dem Schüler deutlich machen. indem sie ihn das zweite Mal dran nimmt?	JA	LP	2.6K1	S. soll besser zuhören	4.7	0.6	2
				2.6K2	*L. demonstriert, dass sie es sieht, wenn S. nicht aufpasst*	*4.1*	*0.9*	*2*
				2.6K3	*Es gelten Regeln in der Klasse. an die man sich halten muss*	*3.7*	*0.9*	*1*
				2.6K4	Interesse an Lernerfolg	3.1	1.4	1
2.7	*Woran kann man erkennen. dass es den SchülerInnen erlaubt ist während der Stunde zu trinken?*	*HP*	*RR*	*2.7K1*	*Trinkbecher.*	*4.8*	*0.4*	*2*
				2.7K2	*Wasserflasche*	*3.8*	*1.4*	*1*
3.1	Geben Sie von den utnerstehenden Aussagen an. ob sie auf dem Abschnitt im Türbereich. bevor die Lerherein mit dem Unterricht anfängt. zutreffen	AP	LP	3.1A	Ein S., der nicht zu Klasse gehört, steht in der Türöffnung	X	X	X
				3.1B	Ein S. kommt zu spät und möchte noch reinkommen	X	X	X
				3.1C	Die L. fragt einen Schüler den Kaugummi auszuspucken	X	X	X
				3.1D	Zwei SuS haben ein persönliches Anliegen, das sie mit der L. besprechen möchten	X	X	X
				3.1E	*Die L. gibt einigen SuS beim Reinkommen die Hand*	*X*	*X*	*X*
				3.1.	Die L. fragt einem S. die Kappe abzunehmen	X	X	X
3.2	*Aus welchen Gründen gibt die Lehrerin sich Mühe. die Ereignisse im Türbereich schnell zu beenden?*	*JA*	*LP*	*3.2K1*	*Zeitsparen*	*3.7*	*1.1*	*2*
				3.2K2	*L. gibt den SuS keine Chance. Störungen zu entwickeln*	*3.8*	*1.1*	*1*
				3.2K3	*L. will unbefugten SuS den Zutritt zum Klassenzimmer verhindern.*	*4.2*	*1.3*	*1*

Note. SSS = situation-specific skills; CM = dimension of classroom management; AP = accuracy of perception; HP, holistic perception; JA = justification of action; CA = classroom arrangement; RR = rules and routines; LP = Low Profile; 3 = Value of third round of expert-ratings.

[a] = Results of 2nd and 3rd round of expert-ratings, X indicates that experts did not rate this criterion.

Italics indicate that an item or Criterion was deleting during the elimination process (see chapter 7.2.4)

Table 22, part V

Item				Criterion		Expert-Rating[a]		
ID	Item	SSS	CM	ID	Criterion	M	SD	3
3.3	Aus welchen Aussagen der Lehrerin kann man schließen. dass die SchülerInnen feste Sitzplätze haben?	HP	CA	3.3K1	Gehe auf Dein Platz	4.7	0.7	1
				3.3K2	Verweis auf geänderten Sitzplan	4.9	0.2	2
3.4	*Wie viele SchülerInnen werden on der Lehrerin namentliche aufgefordert aufzustehen. bevor mit dem Grußaustausch angefangen wird?* *A: keine/r* *B: 1* *C: 2* *D: 3* *E: 4*	*AP*	*RR*	*3.4K1*	*C. 2*	*X*	*X*	*X*
3.5	Welche Gründe könnte die Lehrerin dafür haben. dass sie die SchülerInnen auffordert aufzustehen. bevor sie mit dem Grußaustausch anfängt?	JA	RR	3.5K1	L. bekommt schnelle Übersicht	2.6	1.3	2
				3.5K2	Körperliche Aktivierung	2.9	1.5	1
				3.5K3	Gleichheit der SuS	2.8	1.6	1
				3.5K4	Disziplin/ Respekt	4.4	0.6	1
				3.5K5	L. zeigt Stundenanfang	4.8	0.4	2
				3.5K6	*L. zeigt, Regeln gelten für alle*	*4.2*	*1.1*	*2*
3.6	Beschreiben Sie in Stichworten die verschiedenen Phasen des Grußaustausches. mit dem die Stunde anfängt.	AP	RR	3.6K1	SuS Stehen	4.7	0.5	1
				3.6K2	Gegenseitige Begrüßung.	5.0	0.0	1
				3.6K3	Hinsetzen und singen	4.8	0.4	1
				3.6K4	Wie geht's	4.9	0.3	1
3.7	Benennen Sie in Stichworten die Wirkungen des Grußaustausches auf das Verhalten der SchülerInnen.	HP	RR	3.7K1	SuS sitzen.	4.6	0.5	1
				3.7K2	SuS sind ruhig	3.9	1.2	1
				3.7K3	L. hat Aufmerksamkeit	4.1	0.7	2
				3.7K4	*Die SuS werden körperlich aktiviert.*	*2.9*	*1.4*	*1*
3.8	Nennen Sie die Vorteile der Tischordnung hinsichtlich der Prävention von Unterrichtsstörungen durch die SchülerInnen.	JA	CA	3.8K1	L. kann all SuS sehen	4.1	1.2	2
				3.8K2	SuS können L. / Tafel sehen	4.4	0.7	2
				3.8K3	Schwierige Unterhaltung	4.1	1.0	2
				3.8K4	Gänge frei	3.6	1.1	1

Note. SSS = situation-specific skills; CM = dimension of classroom management; AP = accuracy of perception; HP, holistic perception; JA = justification of action; CA = classroom arrangement; RR = rules and routines; LP = Low Profile; 3 = Value of third round of expert-ratings.

[a] = Results of 2nd and 3rd round of expert-ratings, X indicates that experts did not rate this criterion.

Italics indicate that an item or Criterion was deleting during the elimination process (see chapter 7.2.4)

Table 22, part VI

Item				Criterion		Expert-Rating[a]		
ID	Item	SSS	CM	ID	Criterion	M	SD	3
4.1	Benennen Sie in Stichworten das Verhalten. mit dem der Lehrer eine neue Unterrichtsphase einleitet.	AP	RR	*4.1K1*	*Ausruf*	*4.9*	*0.5*	*2*
				4.1K2	Positionierung	4.1	0.9	2
				4.1K3	Gestik	4.8	0.7	1
				4.1K4	Runterzählen	4.9	0.5	1
4.2	Nennen Sie in Stichworten die Handlungsmaßnahmen. mit denen der Lehrer die SchülerInnen zum Flüstern bewegt.	AP	LP	4.2K1	Verweis auf Unterrichtsroutine	4.5	1.0	2
				4.2K2	L. flüstert selber	4.4	1.0	1
				4.2K3	L. sagt Whisper	4.6	0.8	1
				4.2K4	Gestik	4.5	1.0	2
				4.2K5	Shhh	4.2	1.0	1
				4.2K6	L. Beschreibt gewünscht Schülerverhalten	4.7	0.6	1
4.3	Woran kann man erkennen. dass die Übung zum Textverstehen öfters in der gesehenen Form stattfindet?	HP	RR	4.3K1	L. sagt, wie wir schon oft gemacht haben.	5.0	0.0	2
				4.3K2	L. gibt Minimale Instruktionen.	4.5	0.6	1
				4.3K3	SuS haben keine Nachfragen	4.3	0.8	1
4.4	Welche Vorteile hat die gesehene Arbeitsform hinsichtlich der Vermeidung von potentieller Unterrichtsstörungen durch die SchülerInnen?	JA	RR	4.4K1	SuS schauen in Bücher	3.9	0.9	1
				4.4K2	*L. kann SuS beobachten*	*4.6*	*0.5*	*1*
				4.4K3	*Leise durch Tonband*	*3.2*	*1.2*	*1*
				4.4K4	Fällt auf. wenn SuS sich nicht beteiligen.	4.3	0.7	1
				4.4K5	SuS werden voll beansprucht	3.4	1.1	2
4.5	*Welche der folgenden Aussagen kennzeichnet die Sitzordnung am besten?* *A: In den vordersten Reihen sitzen nur Jungs.* *B: In den vordersten Reihen sitzen nur Mädchen.* *C: In den vordersten Reihen sitzen Jungs und Mädchen gleich gemischt.*	*AP*	*CA*	*4.5K1*	*A. In den vordersten Reihen sitzen nur Jungs*	*X*	*X*	*X*
4.6	Welche Gründe könnte der Lehrer dafür haben. erst zum Tisch der Schülerin zu gehen. bevor er sie tadelt?	JA	CA	4.6K1	Keine Ablenkung anderer SuS	4.5	1.0	2
				4.6K2	Geringe Bloßstellung	3.5	1.3	1
				4.6K3	L. kann die Situation genauer beurteilen	4.5	0.8	1
				4.6K4	Körperliche Nähe erzeugt Dominanz	4.1	1.1	2
4.7	*Was stellt sich im Zwiegespräch als Hauptgrund dafür heraus. dass die Schülerin sich zunächst nicht an der Übung beteiligte?*	*HP*	*LP*	*4.7K1*	*Die S. hatte andere Materialien vor sich*	*4.5*	*1.3*	*2*

Note. SSS = situation-specific skills; CM = dimension of classroom management; AP = accuracy of perception; HP, holistic perception; JA = justification of action; CA = classroom arrangement; RR = rules and routines; LP = Low Profile; 3 = Value of third round of expert-ratings.

[a] = Results of 2nd and 3rd round of expert-ratings, X indicates that experts did not rate this criterion.

Italics indicate that an item or Criterion was deleting during the elimination process (see chapter 7.2.4)

Appendix C: Treatment

The following pages consist of the original instructions of both video-based courses. These instructions include the descriptions of the exercises as handed out to the course participants, the teacher instructions and hand-outs. For the purpose of understanding, teacher-instruction and exercises are placed next to each other in chronological order. The texts containing the instructions for the teachers were not included in the exercise-booklets of the prospective teachers. Boxed texts represent teacher-instructions.

Handreichung Lehre Seminar 2A (Situierte Strategie)

Donnerstag, 4.8.2016

Raum: SH 3.101

Checkliste Materialien:

- TeilnehmerInnenliste
- Tabletliste 2A & 2B
- Kreppbandrollen (2X)
- Permanentmarker
- Parsons Handout (50x)
- Handout zum Einloggen von Vigor
- Übungshefte (25x)
- Klebekärtchen:
    - Gelb (Gestalt)
    - Grün & Rosa (Schemata)
    - Blau (Theorie)
- Pinnwände & Flipcharts
- Farbige Stifte

Einrichten von Vigor 09:15 – 10:00 Uhr (45 Min.)

- Das Seminar wird ‚organisatorisch' eröffnet und die Vigor-Folie an die Leinwand projiziert.
- Kurze Rückmeldung von Herrn Schep: Kann angefangen werden?

Plenum 10:00 – 10:45 Uhr (45 Min.)

Vortrag Prof. Dr. Udo Rauin im Raum 3.104 (45 Min.)

- Alle Studierenden kommen in den Raum 3.104.

Kurze Pause 10:45 – 11:00 Uhr (15 Min.)

Übungen am Vormittag 11:00 – 12:00 Uhr (60 Min.)

Vorstellungsrunde: 11:00 – 11:10 Uhr (10 Min.)

- Jede Person stellt sich kurz in Bezug auf die folgenden Punkte vor: Name, Fächer, Praxiserfahrungen.
- Moderator beschreibt den Verlauf des Tages und des Seminars (Handout):
  - Wir schauen uns Videos an;
  - Wir werden anhand der Videos und Übungen mehr über das Thema Klassenführung lernen.
- Gruppenfindung

Übung 1: Role-Models zu Klassenführung 11:10 – 11:35 Uhr (25 Min.)

a. Rufen Sie sich zwei Ihrer Lehrkräfte in Erinnerung, die Sie aus der Schule kennen: Bei der einen konnte man immer ruhig arbeiten und bei der anderen war es öfters unruhig. Notieren Sie auf ein Papier für sich die Antworten zu folgenden Fragen:
   - Beschreiben Sie für jede Lehrkraft eine typische Situation mit dieser Lehrkraft.
   - Welche Merkmale (Einstellungen, Handlungsweisen, Persönlichkeit) würden Sie in Ihrer späteren Lehrpraxis gerne von dieser Lehrkraft übernehmen und welche würden Sie vermeiden?

a. Diskutieren Sie die Antworten in der Gruppe: Gibt es Ähnlichkeiten/Überlappungen zwischen den einzelnen Gruppenzugehörigen?

b. Jede Person schreibt pro Role-Model das wichtigste Merkmal (einmal vermeiden, einmal übernehmen) auf ein gelbes Kärtchen und hängt es auf die ganz linksstehende Pinnwand.

Ergebnissicherung Übung 1, 11:35 – 12:00 Uhr (25 Min.)

Ziel: Ergebnisse an der ersten Pinnwand (Gestalt, gelb) sichern:

Ablauf:

- Der Moderator gibt die Möglichkeit für Rückfragen zum Verständnis/Erläuterungen.

- Der Moderator fängt an ‚laut zu denken' und zusammen mit den TeilnehmerInnen die Kärtchen zu clustern.
    - Der/die Assistent/in fängt an die gelben Kärtchen zu kategorisieren:
        - LK-Handlung - LK-Persönlichkeit/-Einstellung
        - Übernehmen - Vermeiden
    - Es werden Ähnlichkeiten und Widersprüche expliziert ohne diese zu ‚lösen'.

Mittagpause 12:00 – 13:00 Uhr (60 Min.)

Übungen am Nachmittag 13:00 – 16:00 Uhr (180 Min.)

Übung 2: Do_Video 1, 13:00 – 13:50 Uhr (50 Min.)

a. Schauen Sie sich das Video auf Vigor (Donnerstag/Do_Video 1) einmal ganz an.
b. Schreiben Sie in Stichworten die wichtigsten Ereignisse auf ein Blatt.
c. Schreiben Sie in Stichworten die Antworten zu folgenden Fragen auf (Sie können das Video gern beliebig oft ansehen). Achtung: Belegen Sie Ihre Antworten mit konkreten Beispielen aus dem Video.
    - Welche Herausforderung(en) muss die Lehrkraft in der beobachteten Szene meistern?
    - Wie hat die Lehrkraft auf die Herausforderung reagiert?
    - Welche Handlungsalternativen stehen der Lehrkraft zur Verfügung?
    - Welche Faktoren (maximal fünf) bestimmen Ihrer Meinung nach den Verlauf in der beobachteten Szene?
d. Diskutieren Sie Ihre Antworten in der Gruppe
    - Schreiben Sie die Handlungsalternativen und Einflussfaktoren (jeweils maximal drei), auf die Sie sich einigen können, auf grüne Kärtchen und heften Sie sie an die zweite Pinnwand.

Übung 3: Do_Video 2, 13:50 – 14:40 Uhr (50 Min.)

a. Schauen Sie sich das Video auf Vigor (Donnerstag/Do_Video 2) einmal ganz an.
b. Schreiben Sie in Stichworten die wichtigsten Ereignisse auf.
c. Schreiben Sie in Stichworten die Antworten zu folgenden Fragen auf (Sie können sich das Video gern beliebig oft ansehen). Achtung, belegen Sie Ihre Antworte mit konkreten Beispielen aus dem Video.
    - Welche Herausforderung(en) muss die Lehrkraft in der beobachteten Szene meistern?
    - Wie hat die Lehrkraft auf die Herausforderung reagiert?
    - Welche Handlungsalternativen stehen der LK zur Verfügung?
    - Welche Faktoren (maximal fünf) bestimmen Ihrer Meinung nach den Verlauf in der beobachteten Szene?
d. Diskutieren Ihre Antworten in der Gruppe.
e. Schreiben Sie die Handlungsalternativen und Einflussfaktoren (jeweils maximal drei), auf die Sie sich einigen können, auf ein rosa Kärtchen und heften sie es auch an die zweite Pinnwand.

Ergebnissicherung Übung 2 und 3, 14:40 – 15:15 Uhr (35 Min.)

Ziel: Ergebnisse an der zweiten Pinnwand (Schemata, grün) sichern. Ablauf:

a. Der Moderator gibt die Möglichkeit Fragen zu stellen bzw. stellt selber Rückfragen zum Verständnis und erläutert ggfs.
b. Der Moderator fängt an ‚laut zu denken' und zusammen mit den TeilnehmerInnen die Kärtchen zu clustern.
    - Die Handlungsalternativen werden so viel wie möglich an die Faktoren gekoppelt.

Die Moderationsphase ist vollendet, wenn die TeilnehmerInnen sich einig sind, dass die Faktoren und Handlungsalternativen vollständig dargestellt wurden.

Übung 4: Vergleich der Videos, 15:15 – 15:35 Uhr (25 Min.)

a. Vergleichen Sie beide Videos und die gesicherten Ergebnisse auf den Pinnwänden. Ist es möglich, Gemeinsamkeiten hinsichtlich der Faktoren und Handlungsalternativen zu entdecken?
b. Diskutieren Sie in der Gruppe, inwieweit es möglich ist, aufgrund der heutigen Entdeckungen den Role-Models (Ihren eigenen, und ggf. den von anderen) allgemeine Ratschläge zu geben? Fassen Sie den Ratschlag in Stichworten zusammen, schreiben Sie ihn auf ein blaues Kärtchen und heften Sie es an die dritte Pinnwand.

Ergebnissicherung Übung 4, 15:35 – 16:00 Uhr (25 Min.)

Ziel: Ergebnisse an der dritten Pinnwand sichern.

Ablauf:

a. Der Moderator gibt die Möglichkeit, Fragen zu stellen bzw. stellt selber Rückfragen zum Verständnis und erläutert ggfs. → Moderator fragt z.B.: warum ist der Ratschlag für das Role-Model relevant?
b. Der Moderator fängt an ‚laut zu denken' und zusammen mit den TeilnehmerInnen die Kärtchen zu clustern.
c. Der Moderator versucht zusammen mit den TeilnehmerInnen einen roten Faden zu erstellen zwischen
    a. Role-Model (Gestalt, 1. Pinnwand)
    b. Faktoren und Handlungsalternativen (Schemata, 2. Pinnwand)
    c. Ratschläge (Techniken, 3. Pinnwand)

Abschluss des Tages 16:00 – 17:00 Uhr (60 Min.)

Ziel: kurze Runde

a. Sind wir auf dem richtigen Weg?
b. Fragen und Bemerkungen
c. Morgen mehr Videos!

Fotos der Ergebnisse auf den Pinnwänden machen.

Handreichung Lehre Seminar 2A (Situierte Strategie)

Freitag, 5.8.2016

Raum: SH 3.101

Checkliste Materialien:

- TeilnehmerInnenliste
- Tabletliste
- Krepppapier
- Fair-2-Bogen, A (14) und B (14)
- Papier mit Testinstruktion des FAIR-2-Bogens
- Handout „Routinen“, nach Borich (2015) (25x)
- Handout Modulprüfung (25x)
- Übungshefte für Studierende (25x)
- Pinnwände, Flipcharts, Kärtchen und Reißzwecken

Vigor einrichten und FAIR-2-Test 09:15 – 09:45 Uhr (30 Min.)

- Fair-2 Bögen verteilen
- Fair-2 Testinstruktion vorlesen
- 2x3 Min. Stoppuhr
- Verspätete Personen bitte draußen warten lassen (in Block Lehre nachschreiben lassen).

Kurze Pause 09:45 – 10:00 Uhr (15 Min.)

Übungen am Vormittag, 10:00 – 12:00 Uhr (120 Min.)

Übung 1: Video 1, 10:00 – 10:50 Uhr (50 Min.)

Individuell:

(a) Schauen Sie sich das Video „Fr_Video 1“ in Vigor an und schreiben Sie in Stichworten auf, was Sie gesehen haben. Für die Beantwortung dieser und aller folgenden Fragen können Sie gerne das Video beliebig oft erneut ansehen.

(b) Nennen Sie von den Faktoren, die wir am Donnerstag definiert haben, diejenigen Faktoren, von denen Sie eines oder mehrere Beispiele im Video erkennt haben (minimal zwei, maximal fünf).

(c) Beschreiben Sie die *Handlungen* oder *Aussagen* der Lehrkraft, die die oben beschriebenen Faktoren illustrieren. Bitte beschreiben Sie diese so präzise wie möglich!

(d) Welche Merkmale der beobachteten Lehrkraft, bzw. des beobachteten Unterrichts überlappen mit den Merkmalen Ihres Role-Models?

(e) Diskutieren Sie kurz Ihre Ergebnisse in der Gruppe.

(f) Stellen Sie sich vor, Sie hätten kürzlich in der Schule dieser Lehrerin Ihre Tätigkeit als Lehrkraft aufgenommen und hätten bei ihr hospitiert. Welche drei Fähigkeiten oder Techniken würden Sie anhand des gesehenen Videoabschnittes gerne von ihr übernehmen? Nennen Sie zu jeder Fähigkeit ein im Video gesehenes konkretes Beispiel. Schreiben Sie die Ergebnisse, über die Sie sich einigen können, auf ein Poster. Teilen Sie dafür das Poster zunächst in vier gleiche Teile ein, in dem Sie ein Kreuz malen. Schreiben Sie dann die Ergebnisse dieser Übung rechts oben mit einem farbigen dicken Stift auf ein Poster/Flipchart-Papier.

Übung 2: Video 2, 10:50 – 11:40 Uhr (50 Min.)

(a) Schauen Sie sich das Video „Fr_Video 2“ in Vigor an und schreiben Sie in Stichworten auf, was Sie gesehen haben.

(b) Von den Faktoren, die wir am Donnerstag definiert haben, nennen Sie diejenigen, von denen Sie Beispiele im Video erkannt haben (minimal zwei, maximal fünf).

(c) Beschreiben Sie die Handlungen oder Aussagen der Lehrkraft, die die oben beschriebenen Faktoren illustrieren. Bitte beschreiben Sie diese so präzise wie möglich!

(d) Welche Merkmale der beobachteten Lehrkraft bzw. des beobachteten Unterrichts überlappen mit den Merkmalen Ihres Role-Models?

In der Gruppe:

(e) Welche drei bis fünf Classroom Management-Fähigkeiten/-Techniken könnte diese Lehrkraft verbessern? Schreiben Sie diese mit einem dicken Stift einer anderen Farbe links oben auf das Poster.

(f) Stellen Sie sich vor, die beiden Lehrkräfte wurden im Rahmen einer Coaching-Struktur in der Schule aneinander gekoppelt. Die Lehrkraft aus Video 1 (im Folgenden „Lehrkraft 1“) hat gerade in der Stunde von Lehrkraft aus Video 2 (im Folgenden „Lehrkraft 2“) hospitiert. Bitte stellen Sie das Gespräch zwischen den Lehrkräften auf folgende Weise schematisch dar:

Rechts unten (unter die Fähigkeiten, die Sie gerne übernehmen würden) schreiben Sie die Tipps, welche die Lehrkraft 1 der Lehrkraft 2 geben könnte. Machen Sie das so konkret wie möglich!

Links unten (unter die Verbesserungspunkte von Lehrkraft 2) schreiben Sie in Stichworten eine mögliche Reaktion der Lehrkraft 2 auf die Tipps von Lehrkraft 1. Es geht hier nicht darum eine vermeintliche Ausrede zu erfinden, sondern vielmehr darum, einen ehrlichen und realistischen Grund von Lehrkraft 2 für ihr Handeln zu finden.

Hängen Sie zum Schluss Ihr Poster an die Pinnwand.

Ergebnissicherung im Seminar 11:40 – 12:00 Uhr (20 Min.)

In der Ergebnissicherung werden pro Gruppe ein Tipp und eine Replik besprochen.

Mittagspause 12:00 – 13:00 Uhr (60 Min.)

Übungen am Nachmittag (13:00 – 15:00 Uhr, 120 Min.)

Gruppengespräch: Regeln und Routinen 13:00 – 13:50 Uhr (50 Min.)

Lehr-Lern-Gespräch mit allen Beteiligten im Seminar, in dem folgende Fragen besprochen werden (30 Min.):

- Wiederholung: Was ist die Funktion von Regeln und Routinen im Unterricht?
- Rituale:
  - Wozu Rituale, welche Funktionen hab sie?
- Regeln:
  - Was muss im Unterricht geregelt werden?
  - Auf welche Weise sollten Regeln eingeführt werden?
  - Wie sorgt man dafür, dass die Regeln gefestigt werden?
  - Welche Regeln sind für welche Altersstufe angemessen?
- Routinen:
  - Welche Abläufe lassen sich gut durch Routinen automatisieren?
  - Gibt es Abläufe, die sich dazu weniger eignen?

- Was ist der genaue Unterschied zwischen Ritualen, Regeln und Routinen?

Dann nach dem Lehr-Lerngespräch: Verteilung des Handouts von Borich zum Thema „Routinen".

Aufgabe für Studierende (20 Min.)

- Welche Routinen sind Ihnen bekannt, welche hätten Sie hier nicht erwartet?
- Markieren Sie die Kategorien der Unterrichtsstunde, die Sie gerne im eigenen Unterricht ausprobieren würden.
- Plenum: Diskussion der Ergebnisse

Übung 3: Hospitieren beim Kollegen 13:50 – 14:30 Uhr (40 Min.)

(a) Individuell: Schauen Sie sich das Video „Fr. Video 3" an.

Tun Sie so, als hätten Sie bei diesem Kollegen hospitiert und Sie wären in der Position, den Kollegen zu beraten. Denken Sie an alle Aspekte, die Sie heute und gestern gelernt haben und gehen Sie die relevanten Faktoren (Zeitmanagement, Raumordnung, Regeln und Routinen…) systematisch durch und formulieren Sie pro Faktor einen passenden Ratschlag. Bitte formulieren Sie Ihre Ratschläge als positives Feedback und seien Sie so präzise wie möglich! → Sie können sich das Video gern beliebig oft ansehen.

(b) Diskutieren Sie die Ratschläge (und ihre Formulierung) in Ihrer Gruppe: Gibt es Ratschläge, die Sie machen würden, die bisher noch nicht als „Faktor" auf die Pinnwand gehängt wurden? Notieren Sie Ihre Ergebnisse systematisch, d.h. pro Faktor mindestens einen Ratschlag auf ein Flipchart-Papier und hängen Sie es an eine Pinnwand oder mit einem Magnet auf das White-Board.

Ergebnissicherung 14:30 – 15:00 Uhr (30 Minuten)

Im letzten Abschnitt tauschen Sie sich untereinander über Ihre Poster aus.

Teilen Sie Ihre Gruppe dafür in zwei Teile (z.B. ein Duo und ein Trio). Wenn der Seminarleiter Bescheid sagt, bleibt der eine Teil der Gruppe beim Poster, während der andere Teil das Poster einer anderen Gruppe besucht. Die Gruppenmitglieder, die beim eigenen Poster bleiben, präsentieren die Ergebnisse.

Nach 15 Minuten gibt der Seminarleiter ein Zeichen und wechseln die zwei Teile der Gruppe die Rollen: der Teil, der eine andere Gruppe besucht hat, kehrt zum eigenen Poster zurück und derjenige, der präsentiert hat, besucht ein anderes Poster.

Auslauf, Abschluss und Ausblick 15:00 – 16:00 Uhr (60 Min.)

Information zu Teilnahmenachweis und Modulprüfung 16:00 – 16:30 Uhr (30 Min.)

- Verteilen des Handouts zur Modulprüfung

Handout „Routinen“ (Borich, 2015, p. 94)

Beginning Class Routine

A. Roll call, absentees
B. Tardies
C. Distributing materials

Work Requirement Routine

A. Heading of papers
B. Use of pen or pencil
C. Writing on back of paper
D. Neatness, legibility
E. Incomplete work

Instructional Activity Routine

A. Student attention
B. Obtaining help
C. Student talk
D. Activities to do when work is completed
E. Student movement
F. Bringing materials

Group Activity Routine

A. Expected behavior in group
B. Expected behavior of students out of group
C. Sharing of resources
D. Individual responsibilities
E. Choosing a group leader

Ending Class Routine

A. Putting away supplies, equipment
B. Cleaning up
C. Dismissing class

Interruption Routine

A. Talk among students
B. Turning in work
C. Handing back assignments
D. Getting back assignments
E. Out-of-seat policies

Use of Room/School Area Routine

A. Shared materials
B. Teacher's desk
C. Water fountain, bathroom, pencil sharpener
D. Student desks
E. Learning centers, stations
F. Playground
G. Lunchroom

Assignment Routine

A. Returning in-class assignments
B. Homework assignments
C. Turning in assignments

Checking Assignments in Class Routine

A. Students exchanging papers
B. Making and grading assignments
C. Turning in assignments

Grading Routine

A. Recording grades
B. Grading criteria
C. Contracting with students for grades

Academic Feedback Routine

A. Rewards and incentives
B. Posting student work
C. Communicating with parents
D. Students' record of grades
E. Written comments on assignments

Figure 6.4 Examples of Classroom Routines and Some Topics They May Address

Handreichung Lehre Seminar 2A (Situierte Strategie)

Samstag, 6.8.2016

Raum: SH 3.104

Checkliste Materialien:

- TeilnehmerInnenliste & Tabletliste
- Arbeitshefte
- Handout Bogen Low-Profile nach Borich (2015)
- Handout mit Link zum Evaluationsbogen der LEVEL-Evaluation
- Flipcharts, Pinnwände, Reißzwecken und Stifte

Übungen am Vormittag 09:15 – 11:30 Uhr (135 Min.)

Thematische Einführung & Ausblick auf den Tag 09:15 – 09:25 Uhr (10 Min.)

Übung 1: Ansehen von Video 1, 09:25 – 09:45 Uhr (20 Min.)

Schauen Sie sich das Video „Sa_Video 1“ an und schreiben Sie jede/r für sich die Antworten zu den folgenden Fragen auf:

- Wie stark ist der Einfluss der folgenden Faktoren auf das Gelingen des Unterrichts? Geben Sie Ihre Antwort in Prozentzahlen
  - die Schüler (z.B. weil sie frech sind)
  - das Handeln der Lehrkraft
  - die Situation (ein klassenexterner Faktor, der nicht im Video abgebildet wurde)
- Nennen Sie im Video gesehene Ereignisse, die Ihrer Meinung nach die oben beschriebene Verteilung begründen.

➔ Gruppenbildung (selber bilden)

Übung 2: Beobachtungsbogen 09:45 – 09:50 Uhr (15 Min.)

Sie bekommen jetzt einen Beobachtungsbogen, der bei der Beobachtung von Unterricht - z.B. in der Forschung oder in Videoclubs - eingesetzt wird. Bitte diskutieren Sie diesen Bogen in Ihrer Gruppe anhand folgender Fragen:

- Zu welchem Zweck wird der Bogen eingesetzt?
- Welche Unterrichtsmerkmale bringt er zum Vorschein?
- Wie praktikabel scheint Ihnen das Instrument?
  - Könnten Sie damit (gut) arbeiten?
  - Welche Fragen haben Sie dazu?
- Gibt es sonstige Fragen?

➔ Klärung von Fragen

Übung 3: Unterrichtsanalyse Video 1, 09: 50 – 10:25 Uhr (35 Min.)

(a) Video anschauen
Schauen Sie sich das Video „Sa-Video 1“ in Vigor noch einmal ganz an. Schauen Sie sich das Video jetzt schrittweise in 20 Sekunden Intervallen an, indem Sie alle 20 Sekunden das Video anhalten (00:00-00:20; 00:21-00:40; etc.).

(b) Nutzung des Bogens

Überlegen Sie sich, welche im Bogen beschriebenen Handlungen im gesehenen Abschnitt stattgefunden haben. Setzen Sie einen Strich an der jeweiligen Stelle im Handout (im Sinne einer Strichliste). Wenn nötig schauen Sie sich die Abschnitte des Videos wiederholt an.

Achtung: Codes wie „Change to other activity“ sind nur dann zu vergeben, wenn die Handlung als Classroom Management Technik eingesetzt wird.

(c) Addieren:

Addieren Sie alle Striche einer jeweiligen Code-Kategorie und tragen Sie ein Kreuz in das der Summe entsprechende Kästchen ein.

Übung 4: Unterrichtsanalyse Video 2, 10:25 – 10:55 Uhr (30 Min.)

Führen Sie jetzt die gleichen Schritte für das zweite Video auf einem separaten Blatt Papier aus.

Übung 5: Vergleich und Bewertung 10:55 – 11:15 Uhr (20 Min.)

Diskutieren Sie jetzt die Ergebnisse in Ihrer Gruppe:

(a) Diskutieren Sie das Instrument. Inwieweit hat Ihnen das Arbeiten mit dem Instrument gefallen, fanden Sie es schwierig, einfach?

(b) Vergleichen Sie die bearbeiteten Beobachtungsbögen miteinander. Fangen Sie dabei bei der Handlungskategorie an (Schritt 1). Schauen Sie sich danach die Häufigkeit der gleichcodierten Handlungen an (Schritt 2):

- Was ist in Ihrer Gruppe der Mittelwert pro gesehene Handlung und Handlungskategorie?
- Inwieweit stimmen Ihre Zahlen mit denen der anderen Gruppenmitglieder überein? Wie erklären Sie sich ggfs. eine etwaige Inkonsistenz Ihrer Ergebnisse?

(c) Erstellen Sie jetzt ein Poster, auf dem Sie die Handlungen und Handlungskategorien darstellen (Zeilen) und in zwei Spalten jeweils die in Ihrer Gruppe ermittelten Mittelwerte der beiden Videos nebeneinander auflisten. Heften Sie das Poster an die Pinnwand über Ihre vorherigen Poster.

(d) Welcher Lehrkraft gelingt es effektiver die Klasse zu managen? Auf welche Weise kann man das aus den Ergebnissen ableiten?

Ergebnissicherung Übung 5, 11:15 – 11:30 Uhr (15 Min.)

Der Moderator bespricht im Plenum die Ergebnisse der Gruppen. Studierende werden nach Rückfragen und Erfahrungen gefragt. Lehr-Lern-Gespräch über folgende Frage:

- Wie war die Arbeit mit dem Bogen?
- Inwieweit zeigen die Ergebnisse der Poster, dass die eine Lehrkraft effektiver als die andere Lehrkraft war?
- Welches Merkmal des LehrerInnenhandelns wird mit dem Bogen erfasst?

LEVEL-Evaluationsbogen 11:30 bis 12:00 Uhr (30 Min.)

- Handouts verteilen. Bögen können über das Tablet ausgefüllt werden.

Pause: 12:00 bis 12:30 Uhr (30 Min.)

Übungen am Nachmittag 12:30 bis 13:30 Uhr (60 Min.)

Übung 6: Einbringen Ihres Role-Models 12:30 – 12:40 Uhr (15 Min.):

Denken Sie an die typische Stunde ihres Role-Models zurück.

- Setzen Sie ein farbliches Häkchen hinter die „Codekategorien" auf dem Poster, wenn die Kategorie auf den Lehrstil der Lehrkraft zutrifft:
  - Grün/schwarz: gutes Role-Model
  - Rot: schlechtes Role-Model

Übung 7: Rückblick auf die Erste Übung von heute 12:40 – 12:45 Uhr (15 Min.)

Nachdem Sie die zwei Videos gesehen haben, beantworten Sie bitte nochmal folgende Frage:

Wie stark ist der Einfluss der folgenden Faktoren auf das Gelingen des Unterrichts? Geben Sie Ihre Antwort in Prozentzahlen

- die Schüler (z.B. weil sie frech sind)
- das Handeln der Lehrkraft
- die Situation (ein klassenexterner Faktor, der nicht im Video abgebildet wurde)

Geben Sie an, ob sich nach der Beobachtung der beiden Videos Änderungen ergeben haben.

Ergebnissicherung Übung 6 und 7 sowie Moderation hin zum Thema Geschichte 12:45 – 13:00 Uhr (35 Min.)

Besprechung der Übung 6: Inwieweit stimmt das Role-Model mit den Beobachtungen in den Videos überein?

Besprechung der Übung 7:

- Inwieweit ist die Bravheit der Schüler ein ausschlaggebender Faktor?
- Inwieweit würde Lehrkraft 1 in einer anderen Klasse auch Erfolg haben? Woran hat das gelegen?
- Inwieweit ist Classroom Management bzw. Präsenz greifbar gemacht worden?
- Was ist die Bedingung dafür, dass die Art des Classroom Management von Lehrkraft 1 funktioniert? (→ Moderation in Richtung Vorgeschichte)
- Jetzt haben Studierende wertvolle Beispiele dafür gesehen, wie sie auf Unterrichtsstörungen eingehen können.

Abschluss: Block Lehre 13:30 – 14:00 (30 Min.)

Individuelle Übung 13:30 – 13:45 Uhr (15 Min)

Denken Sie an die letzten drei Tage zurück. Schreiben Sie bitte jede/r für sich Folgendes auf:

1. Allgemein: Das wichtigste für mich, das ich aus dem Seminar mitgenommen habe, ist:

2. Präzise: In der erstfolgenden Praxisphase möchte ich Folgendes direkt in der Klasse ausprobieren/einführen:
3. Wenn das Seminar noch einmal in der Form angeboten werden würde, würde ich Folgendes ändern.

Abschluss in der Gruppe 13:45 – 14:00 Uhr (15 Min)

- Abschlussgespräch, in dem die Antworten der Übung besprochen werden.
- Aufforderung:
    - Lassen Sie sich so oft sie können auf Video aufnehmen:
        - Selbstreflexion
        - Peer-Coaching
        - Video-Clubs

Handout „Low-Profile“ (Borich, 2015, p. 260)

	Times Observed					
	1	2	3	4	5	6
Anticipation	❑	❑	❑	❑	❑	❑
back-and-forth scanning	❑	❑	❑	❑	❑	❑
faster pace	❑	❑	❑	❑	❑	❑
louder voice/high pitch	❑	❑	❑	❑	❑	❑
greater movement	❑	❑	❑	❑	❑	❑
change to other activity	❑	❑	❑	❑	❑	❑
other _____	❑	❑	❑	❑	❑	❑
Deflection (Nonverbal)	❑	❑	❑	❑	❑	❑
moving closer to student	❑	❑	❑	❑	❑	❑
personal eye contact	❑	❑	❑	❑	❑	❑
facial expression	❑	❑	❑	❑	❑	❑
other _____	❑	❑	❑	❑	❑	❑
Deflection (Verbal)	❑	❑	❑	❑	❑	❑
rule reminder (prompting)	❑	❑	❑	❑	❑	❑
name dropping	❑	❑	❑	❑	❑	❑
peer recognition	❑	❑	❑	❑	❑	❑
other _____	❑	❑	❑	❑	❑	❑
Reaction	❑	❑	❑	❑	❑	❑
warning	❑	❑	❑	❑	❑	❑
incentive	❑	❑	❑	❑	❑	❑
loss of privilege	❑	❑	❑	❑	❑	❑
punishment (detention)	❑	❑	❑	❑	❑	❑
other _____	❑	❑	❑	❑	❑	❑

Handreichung Lehre Seminar 2B (Kognitivistische Strategie)

Donnerstag, 4.8.2016

Raum: SH 3.106

Checkliste Materialien:

- TeilnehmerInnenliste
- Übungshefte für Studierende (25x)
- Folie zur Einrichtung von Vigor
- Handout zum Einloggen von Vigor
- Klebekärtchen oder ähnliches, zum Platzieren von Anmerkungen auf den Skizzenpostern
- Pinnwände & Flipcharts
- Farbige Stifte

Einrichten von Vigor 9:15 – 10:00 Uhr (45 Min.)

- Das Seminar wird ‚organisatorisch' eröffnet und die Vigor-Folie an die Leinwand projiziert.
- Es wird auf den kleinen Videotest hingewiesen.
- Kurze Rückmeldung von Herrn Schep: Kann angefangen werden?
- Die Testinstruktion wird vorgelesen und der Test durchgeführt.

Plenum, Vortrag von Prof. Dr. Udo Rauin 10:00 – 10:45 Uhr (45 Min.)

- Alle Studierenden gehen in den anderen Raum (3.104)

Kurze Pause 10:45 – 11:00 Uhr (15 Min.)

Übungen am Vormittag 11:00 – 13:00 Uhr (120 Min.)

Vorstellungsrunde: 11:00 – 11:10 (10 Min.)

- Jede Person stellt sich kurz in Bezug auf die folgenden Punkte vor: Name, Fächer, Praxiserfahrungen.
- Es wird das Handout mit dem Verlauf des Tages und des Seminars verteilt.
- Gruppenfindung

Thema Raumordnung und Methode 11:10-11:25 (15 Min.):

Raumordnung als Dimension von Klassenführung

Vorgehensweise:

- Vormittags: Aufbau eines Klassenzimmers & Wo passiert was?
- Nachmittags: Welche Möglichkeiten bietet der Raum, welche nicht?
    - Beteiligungsmöglichkeiten
    - Beteiligungsbarrieren
- Bewertung der Situation und Suchen nach Alternativen

Instruktion zur Vormittagsübung:

Bearbeiten Sie nacheinander die folgenden Aufgaben gemeinsam in Ihrer Kleingruppe. Ziel ist das gemeinsame Erstellen von zwei Postern, auf denen die Ergebnisse der einzelnen Arbeitsschritte festgehalten werden und welche am Ende zum Vergleich der verschiedenen Gruppenergebnisse dienen.

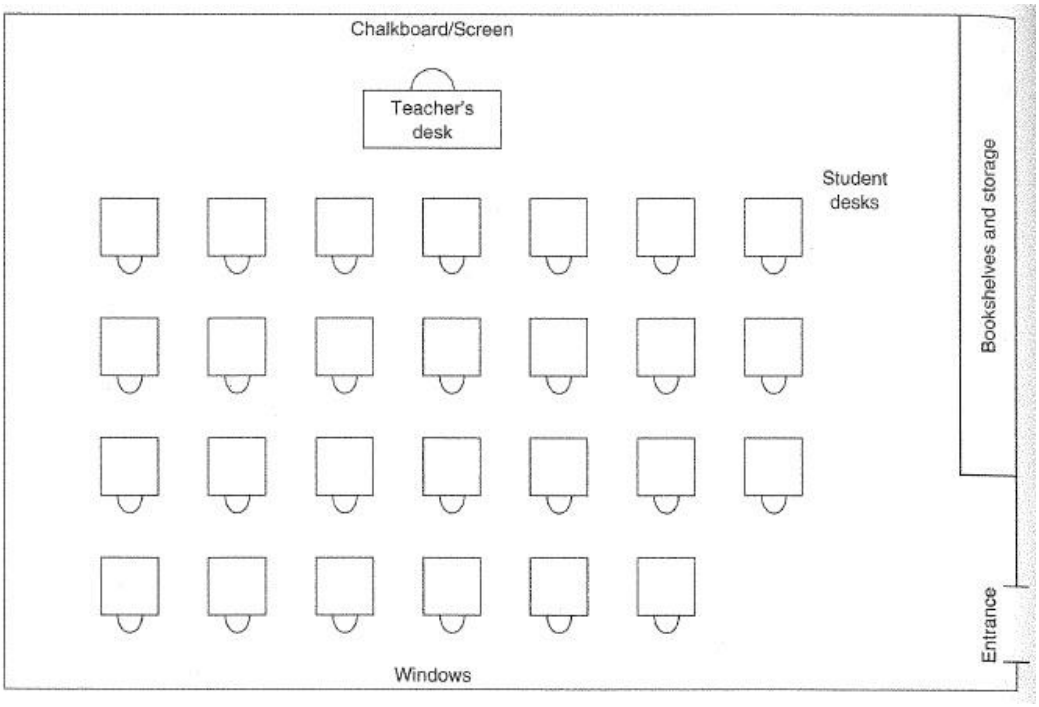
1a) Der Aufbau eines Klassenraums 11:25 – 11:45 Uhr (20 Min.)

Zuerst soll ein Klassenraum in seinem „Ruhezustand" schematisch beschrieben werden. Schauen Sie sich im Video „Do_Video 1"[10] den Aufbau des Klassenraums an und fertigen Sie

[10] Hecht (2008) Selbsttätigkeit im Unterricht: Empirische Untersuchungen in Deutschland und Kanada zur Paradoxie pädagogischen Handelns

auf dem Poster eine Skizze des Raumes aus der Vogelperspektive an. Ähnlich wie in der Beispielskizze unten sollen der Grundriss des Raums, Mobiliar, Geräte und die Positionen der beteiligten Personen schematisch dargestellt werden.

Für die Begutachtung des Raums können Sie jede beliebige Stelle im Video nutzen, weil die Kameraperspektive sich nicht ändert. Gerne können Sie kleine Vorskizzen anfertigen.



Beispielhafte Skizze eines Klassenraums (Borich, (1994) Observation Skills. S. 136)

1b) Wo passiert was? 11:45 – 12:15 Uhr (30 Min.)

Im nächsten Schritt soll betrachtet werden, was an welcher Stelle im Raum passiert:

(a) Beobachten Sie dazu die Bewegungen der Lehrkraft im Raum und zeichnen Sie mit grüner Farbe die Laufwege und Bewegungszonen der Lehrkraft in die Raumskizze ein.

(b) Wenn Sie den Blick auf die Schüler/innen richten: Auf welche Punkte im Raum richten die Schüler/innen in dieser Phase hauptsächlich ihre Aufmerksamkeit? Versehen Sie die entsprechenden Gegenstände oder Personen in Ihrer Raumskizze mit blauen Kreisen.

(c) Nun zu den beobachtbaren Arbeitsformen: Welche Arbeitsformen können Sie im Video erkennen? (z.B. Lehrervortrag, Klassengespräch, Einzelarbeit, Spiel, Gruppenarbeit, Partnerarbeit, etc.)

Notieren Sie die gefundenen Unterrichtsformen als lose Begriffssammlung am Rande Ihrer Raumskizze!

→ Hängen Sie das Poster abschließend an eine der bereitstehenden Pinnwände!

2a) Der Aufbau eines zweiten Klassenraums 12:15 – 12: 35 Uhr (20 Min.)

Nun geht es darum, die vorherigen Analyseschritte auf eine andere Unterrichtssituation anzuwenden.

Fertigen Sie dazu ebenfalls für das Video „Do_Video 2“[11] eine solche Skizze zum Aufbau des Klassenraums an wie in Arbeitsschritt 1a. Verwenden Sie dafür ein neues Poster.

Achtung: Im Video gibt es zwei unterschiedliche Weisen, wie der Klassenraum geordnet wurde. Fertigen Sie für beide Weisen jeweils eine Skizze an. Bearbeiten Sie folgende Aufgaben bitte für beide Raumordnungen.

2b) Auch hier: Wo passiert was? 12:25 – 13:00 Uhr (25 Min.)

Halten Sie auch für diese Unterrichtsstunde fest, was an welcher Stelle im Raum passiert:

(a) Beobachten Sie dazu – in beiden Raumordnungen - die Bewegungen der Lehrkraft im Raum und zeichnen Sie mit grüner Farbe die Laufwege und Bewegungszonen der Lehrkraft in die Raumskizze ein.

(b) Der Blick auf die Schüler/innen: Versehen Sie die Gegenstände oder Personen, auf die die Schüler/innen ihre Aufmerksamkeit richten mit blauen Kreisen.

(c) Welche Arbeitsformen können Sie im Video erkennen? (z.B. Lehrervortrag, Klassengespräch, Einzelarbeit, Spiel, Gruppenarbeit, Partnerarbeit, etc.)

Notieren Sie die gefundenen Unterrichtsformen als lose Begriffssammlung am Rande Ihrer Raumskizze!

→ Hängen Sie das Poster abschließend an eine der bereitstehenden Pinnwände!Mittagspause 13:00 – 14:00 Uhr (60 Min.)

Übungen am Nachmittag 14:00 – 16:00 Uhr (120 Min.)

3. Welche Möglichkeiten bietet der Raum – und welche nicht?

Jetzt, wo wir zu beiden Videos genau wissen, wer und was sich an welchen Stellen im Raum befindet und wann was passiert, geht es darum, zu betrachten, welche Konsequenzen das für den Unterricht hat.

3a) Beteiligungsmöglichkeiten 14:00 – 14:20 Uhr (20 Min.)

Wenn Sie im ersten Video („Do_Video 1“) die Kombination aus Raumskizze und gesehenen Arbeitsformen betrachten: Welche Möglichkeiten haben die Schüler/innen, sich an den Unterrichtsinhalten zu beteiligen?

Schauen Sie sich dazu Ihre Ergebnisse vom Vormittag auf dem Poster an und ergänzen Sie diese durch kurze Eindrücke aus dem Video.

Notieren Sie stichpunktartig, durch welche Tätigkeiten es den Schüler/innen hier möglich ist, sich mit den Unterrichtsinhalten zu beschäftigen:

Video „Do_Video 1“

Unterrichtliche Arbeitsform	**Position der Lehrkraft**	**Beteiligung der Schüler/innen**

[11] Quelle: Projekt UDiKom (Helmke, Helmke, Lenske, Pham, Praetorius, Schrader & Ade-Thurow), http://www.unterrichtsdiagnostik.info/video/

3b. Beteiligungsbarrieren 14:20 – 14:30 Uhr (10 Min.)

Fallen Ihnen mögliche Bewegungs- oder Wahrnehmungsbarrieren im Raum auf, die die Schüler/innen, aber auch den Lehrer möglicherweise daran hindern, sich so im Raum zu bewegen, wie es die Arbeitsformen verlangen, bzw. Ihre Aufmerksamkeit so auszurichten, wie es nötig ist? Falls ja, markieren Sie diese Zonen auf Ihrem Poster mit roten Strichen.

4. Anwendung auf das zweite Video:

Wiederholen Sie nun diesen Analyseschritt, indem Sie ihn auf das Video „Do_Video 2“ anwenden!

4a. Beteiligungsmöglichkeiten 14:30 – 14:50 Uhr (20 Min.)

Video „Do_Video_2“

Unterrichtliche Arbeitsform	**Position der Lehrkraft**	**Beteiligung der Schüler/innen**

4b. Beteiligungsbarrieren 14:50 – 15:00 Uhr (10 Min.)

Markieren Sie auch hier mögliche Bewegungs- oder Wahrnehmungsbarrieren im Raum durch rote Striche in den entsprechenden Zonen auf der Raumskizze.

5. Bewertung der Situation und Suche nach Alternativen 15:00 – 15:30 Uhr (30 Min.)

Wenn Sie alles, was Sie nun über diese beiden Unterrichtsstunden herausgefunden haben, zusammen betrachten:

(a) Wie würden Sie jeweils die Eignung der Raumordnung zu den vollzogenen Arbeitsformen bewerten? Einigen Sie sich in Ihrer Gruppe zu jedem der beiden Videos auf ein Urteil auf der Schulnotenskala und kleben Sie die Note mit einem Klebekärtchen mit der Überschrift „Raumordnung/Arbeitsformen: [Note]“ an eine Ecke der jeweiligen Raumskizze.

Berücksichtigen Sie in Ihrer Bewertung bewusst den Vergleich zwischen den beiden Unterrichtsstunden!

Tipp: Wenn es Ihrer gemeinsamen Entscheidungsfindung hilft, können Sie gerne noch einmal in die Videos schauen!

(b) Könnte man die jeweilige Raumordnung womöglich optimieren, um auch über die gesehene Unterrichtsstunde hinaus besser auf verschiedene Unterrichtsformen eingestellt zu sein? Wenn ja, wie? Notieren Sie entsprechende Veränderungsvorschläge in Kurzform auf Klebekärtchen (z.B. „Sitzordnung U“) und kleben Sie sie an die Stellen auf der Raumskizze, wo Sie etwas verändern würden.

5. Ergebnissicherung 15:30 – 16:00 Uhr (30 Min.)

Marktplatzartige Ergebnisschau kommentiert durch den Dozenten

Abschluss des Tages 16:00 – 17:00 Uhr (60 Minuten)

Ziel: kurze Runde:

- Fragen und Bemerkungen
- Ausblick auf morgen

Einsammeln:

- Tablets
- Tablets zählen

Fotos der Ergebnisse auf den Pinnwänden machen

Handreichung Lehre Seminar 2B (Kognitivistische Strategie)

Freitag, 5.8.2016

Raum: SH 3.106

Checkliste Materialien:

- TeilnehmerInnenliste
- FAIR-2-Bogen, A (14) und B (14)
- Papier mit Testinstruktion des FAIR-2-Bogens
- Handout „Routinen“ nach Borich (2015) (25x)
- Handout Modulprüfung (25x)
- Übungshefte für Studierende (25x)
- Pinnwände, Flipcharts, Kärtchen und Reißzwecken.

Vigor einrichten und FAIR-2-Test 09:15 – 09:45 Uhr (30 Min.)

- Fair-2 Bögen verteilen
- Fair-2 Testinstruktion vorlesen
- 2x3 Min. Stoppuhr
- Verspätete Personen bitte draußen warten lassen (in Block Lehre nachschreiben lassen).

Kurze Pause 09:45 – 10:00 Uhr (15 Min.)

Übungen am Vormittag 10:00 – 12:00 Uhr (120 Min.)

Tagesablauf:

- Theoretische Vertiefung
- Videos analysieren
- Vergleichen und Bewerten
- Ratschläge für Verbesserungen

Thematische Einführung & Ausblick auf den Tag, 10:00 – 10:20 Uhr (20 Min.)

Erläuterungen zur Rolle von Regeln und Routinen im Unterricht

- Es gibt viele wiederkehrende Abläufe im Unterricht (z.B. Stundenbeginn/-ende, Phasenübergänge, Materialausgabe, Wortmeldungen, Tests, Organisatorisches wie Tafel wischen etc.).
- Gefahr des Verlusts von Unterrichtszeit, wenn diese Abläufe jedes Mal von neuem und jedes Mal anders organisiert und den Studierenden erklärt werden müssten → Notwendigkeit von Routinen.
- Regeln haben eine ähnliche Funktion: Sie werden eingeführt um schnell arbeiten zu können, ohne viel über die Bedingungen diskutieren und verhandeln zu müssen. Regeln sind allerdings sehr präzise, an spezifischem SchülerInnenverhalten orientierte Prinzipien:
    - Erst die Hand heben, bevor man sich zu Wort meldet;
    - Wenn die Lehrkraft redet, sind die SchülerInnen aufmerksam;
    - Im Klassenzimmer wird nicht gegessen.
- Hier wird klar: Regeln haben noch eine Funktion, die über die Zeitersparnis hinausgeht: Die Vermeidung von Konflikten, weil SchülerInnen wissen, woran sie sind.
- Unterscheidung:
    - Regeln beziehen sich auf Normen
    - Routinen sind ein Bündel von spezifischen Regeln, die sich auf bestimmte Zeitpunkte, Orte oder Unterrichtsaspekte (z.B. Gruppenarbeit, Phasenübergänge) beziehen (Borich, 2015, S. 93).
- Für sowohl Regel als auch Routine gilt, dass sie etabliert und aufrecht erhaltet werden müssen → Achtung: Wenn der Unterricht gut funktioniert, sind sie schwierig zu entdecken.
- Heute setzen wir uns nun mit bestimmten Aspekten davon auseinander!

Übung 1: Analyse Stundeneinstieg Video 1, 10:20 – 11:05 Uhr (45 Min.)

1. Die Chronik des Unterrichtsbeginns

Das Hauptziel der Etablierung und Einhaltung von Regeln und Routinen ist die Minimierung von Zeitverlusten im Unterrichtsverlauf bzw. die Maximierung der Lernzeit. Ein wichtiger Punkt dafür ist der Unterrichtsbeginn. Um dies zu verdeutlichen, wollen wir uns im Folgenden Situationen des Unterrichtseinstiegs einmal im Detail anschauen.

(a) Erstellen einer Chronik

Schauen Sie sich das Video „**Fr_Video 1“** an und protokollieren Sie anschließend anhand der beiliegenden Mustertabelle die Chronologie der Handlungen der Unterrichtsbeteiligten auf ein Flip-Chart-Papier (Zeitpunkt & beobachtetes Verhalten, s.a. Beispieleintrag). Beziehen Sie sich besonders auf die nachfolgend aufgeführten Handlungen und schauen Sie sich das Video dazu am besten zwei- oder dreimal an!

Lehrkraft:

- Anwesenheit im Raum (Lehrkraft betritt oder verlässt den Raum)
- Hand- oder Körpersignale
- pseudo-verbale Signale („Psch!“)
- verbale Aufforderungen
- Ermahnungen
- Sonstige Äußerungen

Schüler/innen:

- Mindestens ein/e Schüler/in tut, was die Schüler/innen aktuell tun sollen
- (Fast) Alle Schüler/innen tun, was sie tun sollen
- Störung/Verstoß gegen die aktuell gültige Verhaltenskonvention
- Schüler/innen übernehmen teilweise die Lehrerrolle/unterstützen die Lehrkraft („Sei doch mal ruhig jetzt!“)
- Sonstiges

Zeit	**Handlungen...**	
	...der Lehrkraft (L)	**...der Schüler/innen (S)**
0:24	L betritt den Raum	
01:13	„Lasst uns anfangen!“ + erhobene Hand	
01:18		Erste Tischgruppe ist bereit
02:05		Fast alle S sind bereit

(b) Benötigte Zeit einschätzen

Sehen Sie sich die von Ihnen zusammengetragene Chronik an und prüfen Sie anhand Ihrer Beobachtungen, wie viel Zeit die Lehrkraft für den Unterrichtsbeginn benötigt hat. Markieren Sie dazu in der Tabelle die aus Ihrer Sicht erste erkennbare Handlung der Lehrkraft zur Herstellung der Unterrichtsöffentlichkeit einerseits, und den Zeitpunkt, an dem die Unterrichtsöffentlichkeit (weitgehend oder vollständig) hergestellt war andererseits, und notieren Sie am Rand der Tabelle die Zeitdifferenz.

(c) Regeln und Routinen identifizieren

Wo sehen Sie in den beobachteten Abläufen Belege dafür bzw. Hinweise darauf, dass

- **+** Regeln und/oder Routinen etabliert sind und funktionieren,
- **0** Regeln und/oder Routinen existieren, aber nicht funktionieren,
- **-** keine Regeln und/oder Routinen existieren.

Markieren Sie dazu entsprechende Stellen in der Chronik mit den Zeichen **+/0/-** !

(d) Belegen anhand von Beispielen

Woran machen Sie diese Einschätzungen fest? Woran erkennen Sie in dem Video die (Nicht-) Existenz von Regeln oder die Wirksamkeit von Routinen? Notieren Sie neben jeweils einer der „+/0/-“ Markierungen ein entsprechendes Stichwort!

Übung 2: Analyse des Stundeneinstiegs, Video 2, 11:05 – 11:50 Uhr (45 Min.)

Erstellen Sie eine weitere Chronik, in dem Sie die Arbeitsschritte von Übung 1 für das Video „Fr. Video_2“ wiederholen.

Übung 3: Vergleich und Bewertung 11:50 – 12:00 Uhr (10 Min.)

Bewertung: Jetzt, wo Sie detaillierte Einblicke in den Ablauf des Stundenbeginns von zwei unterschiedlichen Unterrichtsstunden gewonnen haben, vergleichen Sie die beiden „Chroniken“ und bewerten Sie sie: Wo läuft es reibungsloser? Wo sehen Sie die bessere Zeitnutzung?

Einigen Sie sich in Ihrer Gruppe zu jedem der beiden Videos auf ein Urteil auf der Schulnotenskala und kleben Sie die Note mit einem Klebekärtchen mit der Überschrift „Zeitmanagement: [Note]“ an eine Ecke der jeweiligen Tabelle.

Berücksichtigen Sie in Ihrer Bewertung bewusst den Vergleich zwischen den beiden Unterrichtsstunden!

Tipp: Wenn es Ihrer gemeinsamen Entscheidungsfindung hilft, können Sie gerne noch einmal in die Videos schauen!

Mittagspause 12:00 bis 13:00 Uhr (60 Min.)

<u>Übungen am Nachmittag 13:00 – bis 15:00 Uhr (120 Min.)</u>

Übung 4.: Routinen 13:00 – 13:20 Uhr (20 Min.)

Markieren Sie in der Tabelle, die Sie als Handout bekommen haben, welche Routinen Sie im Video gesehen haben. Welche der beschriebenen Routinen hätten Ihrer Meinung nach den Verlauf des Unterrichts noch verbessert? Markieren Sie diese Routinen auf dem Handout.

Heften Sie abschließend Ihre beiden Chroniken und das Blatt mit den Notizen aus Übung 4 an die Pinnwand.

Übung 5: Hospitieren bei einem Kollegen 13:20 – 14:10 Uhr (50 Min.)

(a) Individuell: Schauen Sie sich das Video „Fr. Video 3“ an.

Tun Sie so, als hätten Sie bei diesem Kollegen hospitiert. Denken Sie an alle Dinge, die Sie heute und gestern gelernt haben: Was würden Sie dem Kollegen hinsichtlich folgender Aspekte raten:

- Ordnung seines Klassenzimmers
- Nutzung von Zeit; d.h. das ‚routinifizieren‘ bestimmter Unterrichtsabläufe und einüben von Regeln.
- Ggfs. andere Aspekte

Bitte formulieren Sie Ihre Ratschläge als positives Feedback und seien Sie so präzise wie möglich! → Sie können sich das Video gern beliebig oft ansehen.

(b) Diskutieren Sie die Ratschläge in Ihrer Gruppe, schreiben Sie diejenigen auf, über die sie sich einigen können und heften Sie das Blatt an Ihre Pinnwand.

Übung zur Ergebnissicherung am Nachmittag 14:10 – 15:00 Uhr (50 Min.)

Jetzt tauschen Sie sich über die Ergebnisse auf Ihren Postern aus. Teilen Sie Ihre Gruppe dafür in zwei Untergruppen (z.B. ein Duo und ein Trio). Der eine Teil steht im äußeren Kreis des Kugellagers beim Poster. Der andere Teil steht im Inneren und dreht sich alle drei Minuten ein Poster weiter nach rechts. Nach 20 Minuten gibt der Seminarleiter ein Zeichen und wechseln Sie die Rollen: diejenigen, die außen standen, wechseln nach innen und umgekehrt. Wieder wechseln die Personen im Inneren des Kugellagers alle drei Minuten ihre Position.

Abschluss und Ausblick 15:00 – 16:00 Uhr (60 Min.)

Information zu Modulprüfung 16:00 – 17:00 Uhr (60 Min.)

- Verteilen der Handouts mit Informationen zu Modulprüfung
- Beantworten von Fragen
- Materialien einsammeln
- Fotos der Pinnwände machen

Handout „Routinen“ (Borich, 2015, p. 94)

Beginning Class Routine

A. Roll call, absentees
B. Tardies
C. Distributing materials

Work Requirement Routine

A. Heading of papers
B. Use of pen or pencil
C. Writing on back of paper
D. Neatness, legibility
E. Incomplete work

Instructional Activity Routine

A. Student attention
B. Obtaining help
C. Student talk
D. Activities to do when work is completed
E. Student movement
F. Bringing materials

Group Activity Routine

A. Expected behavior in group
B. Expected behavior of students out of group
C. Sharing of resources
D. Individual responsibilities
E. Choosing a group leader

Ending Class Routine

A. Putting away supplies, equipment
B. Cleaning up
C. Dismissing class

Interruption Routine

A. Talk among students
B. Turning in work
C. Handing back assignments
D. Getting back assignments
E. Out-of-seat policies

Use of Room/School Area Routine

A. Shared materials
B. Teacher's desk
C. Water fountain, bathroom, pencil sharpener
D. Student desks
E. Learning centers, stations
F. Playground
G. Lunchroom

Assignment Routine

A. Returning in-class assignments
B. Homework assignments
C. Turning in assignments

Checking Assignments in Class Routine

A. Students exchanging papers
B. Making and grading assignments
C. Turning in assignments

Grading Routine

A. Recording grades
B. Grading criteria
C. Contracting with students for grades

Academic Feedback Routine

A. Rewards and incentives
B. Posting student work
C. Communicating with parents
D. Students' record of grades
E. Written comments on assignments

Figure 6.4 Examples of Classroom Routines and Some Topics They May Address

Handreichung Lehre Seminar 2B (Kognitivistische Strategie)

Samstag, 6.8.2016

Raum: SH 3.106

Checkliste Materialien:

- TeilnehmerInnenliste
- Arbeitshefe
- Handout Bogen Low-Profile nach Borich (2015)
- Handout mit Link zum Evaluationsbogen der LEVEL-Evaluation
- Flipcharts, Pinnwände, Reißzwecken und Stifte

Übungen am Vormittag (09:15 – 11:30 Uhr, 135 Min.)

Start 09:15 – 09:20 Uhr (5 Min.):

- Willkommen und Ausblick auf den Tagesablauf

Vortrag zu Low-Profile 09:20 – 09:45 Uhr (25 Min.)

- Idee Low-Profile
- Einführung des Handouts
- Einführung des Rater-Verfahrens → Intervallkodierung (ggf. Inter-Rater-Reliabilität?)

Übung 1: Analyse Low Profile Video 1, 09:45 – 10:20 Uhr (35 Min.)

(a) Video anschauen
Schauen Sie sich das Video „Sa-Video 1" in Vigor einmal ganz an. Schauen Sie sich dann das Video noch einmal, aber jetzt schrittweise in 20 Sekunden Intervallen an, indem Sie alle 20 Sekunden das Video anhalten (00:00-00:20; 00:21-00:40; etc.).

(b) Kodieren

Überlegen Sie sich, wie bestimmte Lehrkrafthandlungen vollzogen wurden und welcher „Code" am besten zu ihnen passen würde. Machen Sie einen Strich an der jeweiligen Stelle im Handout (im Sinne einer Strichliste). Wenn nötig schauen Sie sich die Abschnitte des Videos wiederholt an.

Achtung: Codes wie „Change to other activity" sind nur dann zu vergeben, wenn die Handlung als Classroom Management Technik eingesetzt wird.

(c) Addieren

Addieren Sie alle Striche einer jeweiligen Code-Kategorie und tragen Sie ein Kreuz in das der Summe entsprechende Kästchen ein.

Übung 2: Analyse Low Profile Video 2, 10:20 – 10:55 Uhr (35 Min.)

Führen Sie jetzt die gleichen Schritte für das zweite Video auf dem zweiten Handout aus.

Übung 3: Vergleich und Bewertung 10:55 – 11:15 Uhr (20 Min.)

Diskutieren Sie jetzt die Ergebnisse in Ihrer Gruppe:

(a) Diskutieren Sie das Instrument. Inwieweit hat Ihnen das Codieren mit dem Instrument gefallen? Fanden Sie es schwierig, einfach?

(b) Vergleichen Sie die bearbeiteten Beobachtungsbögen miteinander. Fangen Sie dabei bei der Code-Kategorie an (Schritt 1). Schauen Sie sich danach die Zahl der vergebenen *Codes* an (Schritt 2):

- Was ist in Ihrer Gruppe der Mittelwert pro Code und Code-Kategorie?
- Inwieweit stimmen Ihre Zahlen mit denen der anderen Gruppenmitglieder überein? Wie erklären Sie sich ggfs. eine etwaige Inkonsistenz Ihrer Ergebnisse?

(c) Erstellen Sie jetzt ein Poster, auf dem Sie in der erste Spalte die Codes und Code-Kategorien darstellen und in der zweiten und dritten Spalten die Mittelwerte der beiden Videos aus Ihrer Gruppe auflisten. Heften Sie das Poster an die Pinnwand über Ihr vorheriges Poster.

(d) Welcher Lehrkraft gelingt es effektiver, die Klasse im Sinne von Borichs Low-Profile zu managen?

Ergebnissicherung Übung 3, 11:15 – 11:30 Uhr (15 Min.)

Der Moderator bespricht im Plenum die Ergebnisse der Gruppen. Studierende werden nach Rückfragen und Erfahrungen gefragt. Lehr-Lern-Gespräch über folgende Frage:

- Wie war die Arbeit mit dem Bogen?
- Inwieweit zeigen die Ergebnisse der Poster, dass eine Lehrkraft möglicherweise effektiver als die andere Lehrkraft war?

Bögen Level-Evaluation 11:30 – 12:00 Uhr (30 Min.)

- Handouts austeilen. Die Bögen können über das Tablet ausgefüllt werden.

Kurze Pause 12:00 bis 12:30 Uhr (30 Min.)

Übungen am Nachmittag 12:30 – 13:30 Uhr (60 Min.)

Übung 4: Interpretation 12:30 – 13:00 Uhr (30 Minuten)

Bereiten Sie individuell Antworten auf die folgenden Fragen vor:

(a) Auf welches typische SchülerInnenhandeln ist Low-Profile Management eine passende Reaktion? Auf welches nicht?

(b) Haben Sie Abschnitte in Video 2 gesehen, wo High-Profile Management größere Störungen auslöst? Falls ja, an welcher Stelle und wie würden Sie das Lehrerhandeln beschreiben?

(c) Was würde passieren, wenn beide Lehrkräfte plötzlich die Klasse wechseln würden? Mit anderen Worten: Inwieweit sind die Schwierigkeiten in der Stunde auf schwierige, bzw. brave SchülerInnen zurück zu führen? Belegen Sie Ihre Antwort mit einem konkreten Beispiel.

(d) Inwieweit würde es der Lehrkraft aus dem ersten Video etwas bringen, wenn sie plötzlich Low-Profile CM durchführen würde?

Plenumsgespräch 13:00 – 13:30 Uhr (30 Min.)

Lehr-Lerngespräch, in dem die Lehrperson die Antworten der Studierenden abfragt und das Gespräch in Richtung des Faktors ‚Vorgeschichte' moderiert.

Abschluss Block Lehre 13:30 – 14:00 Uhr (30 Min.)

Zusammenfassung 13:30 – 13:45 Uhr (15 Min)

Denken Sie an die letzten drei Tage zurück. Schreiben Sie bitte jede/r für sich Folgendes auf:

1. Allgemein: Das wichtigste für mich, das ich aus dem Seminar mitgenommen habe, ist:
2. Präzise: In der erstfolgenden Praxisphase möchte ich Folgendes direkt in der Klasse ausprobieren/einführen:
3. Wenn das Seminar noch einmal in der Form angeboten werden würde, würde ich Folgendes ändern.

Abschluss im Plenum 13:45 – 14:00 Uhr (15 Min.)

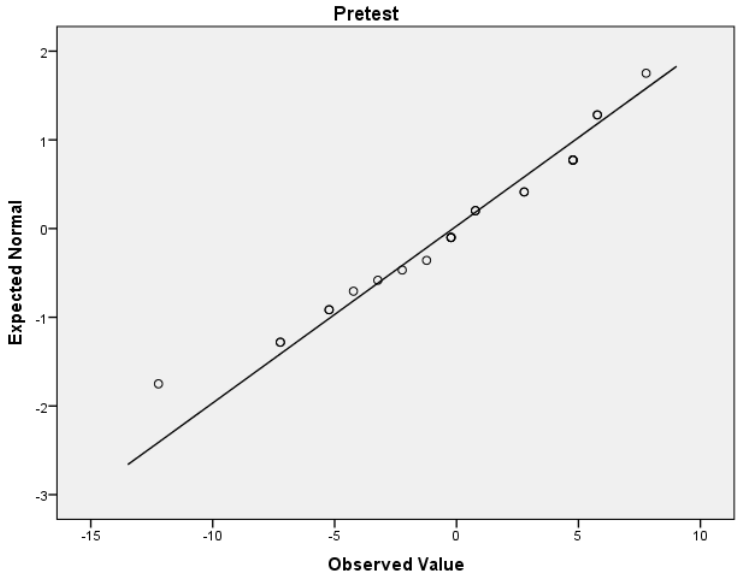
- Abschlussgespräch, in dem die Antworten der Übung besprochen werden.
- Aufforderung:
    - Lassen Sie sich so oft sie können auf Video aufnehmen:
        - Selbstreflexion
        - Peer-Coaching
        - Video-Clubs

Handout „Low-Profile“ (Borich, 2015, p. 260)

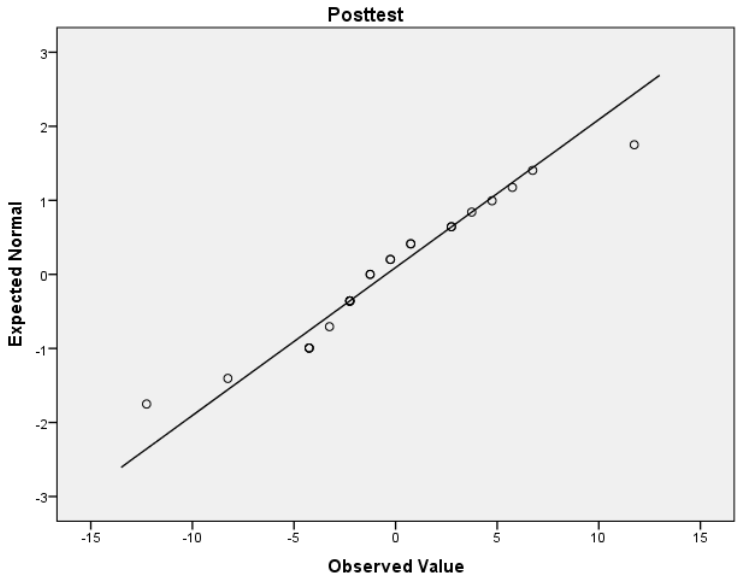
	Times Observed					
	1	2	3	4	5	6
Anticipation	❑	❑	❑	❑	❑	❑
back-and-forth scanning	❑	❑	❑	❑	❑	❑
faster pace	❑	❑	❑	❑	❑	❑
louder voice/high pitch	❑	❑	❑	❑	❑	❑
greater movement	❑	❑	❑	❑	❑	❑
change to other activity	❑	❑	❑	❑	❑	❑
other _____	❑	❑	❑	❑	❑	❑
Deflection (Nonverbal)	❑	❑	❑	❑	❑	❑
moving closer to student	❑	❑	❑	❑	❑	❑
personal eye contact	❑	❑	❑	❑	❑	❑
facial expression	❑	❑	❑	❑	❑	❑
other _____	❑	❑	❑	❑	❑	❑
Deflection (Verbal)	❑	❑	❑	❑	❑	❑
rule reminder (prompting)	❑	❑	❑	❑	❑	❑
name dropping	❑	❑	❑	❑	❑	❑
peer recognition	❑	❑	❑	❑	❑	❑
other _____	❑	❑	❑	❑	❑	❑
Reaction	❑	❑	❑	❑	❑	❑
warning	❑	❑	❑	❑	❑	❑
incentive	❑	❑	❑	❑	❑	❑
loss of privilege	❑	❑	❑	❑	❑	❑
punishment (detention)	❑	❑	❑	❑	❑	❑
other _____	❑	❑	❑	❑	❑	❑

Appendix D: Q-Q-Plots on all Measurements of CMC

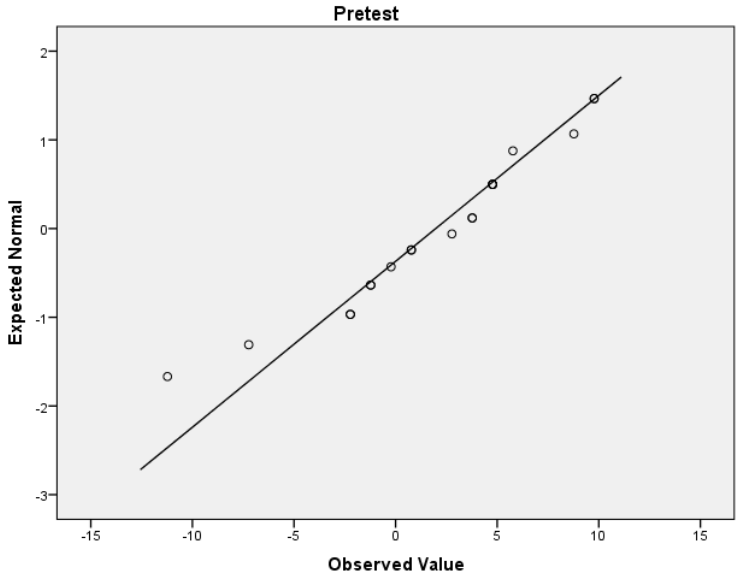
Normal Q-Q Plots for Residuals in Group 1Sit. at Pretest



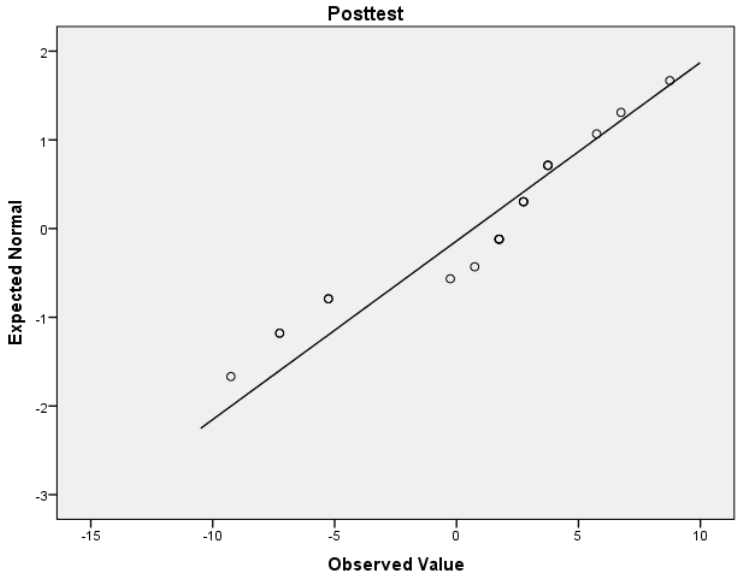
Normal Q-Q Plots for Residuals in Group 1Sit. at Posttest



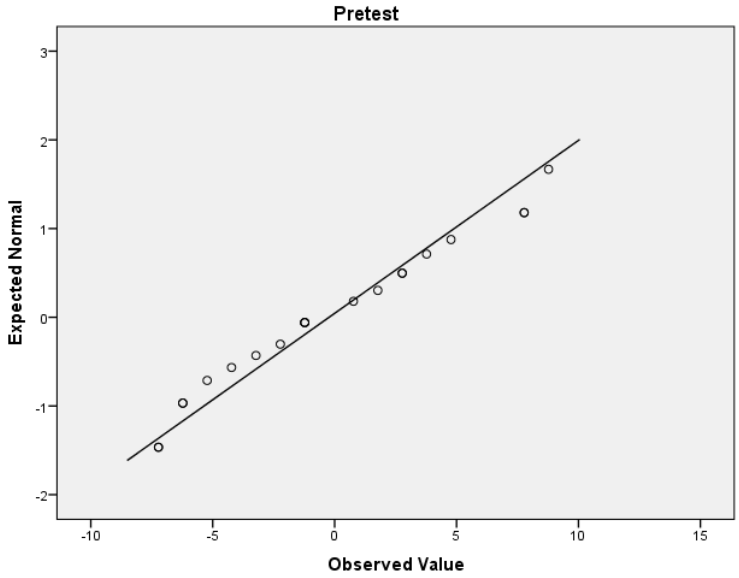
Normal Q-Q Plots for Residuals in Group 1Cog. at Pretest



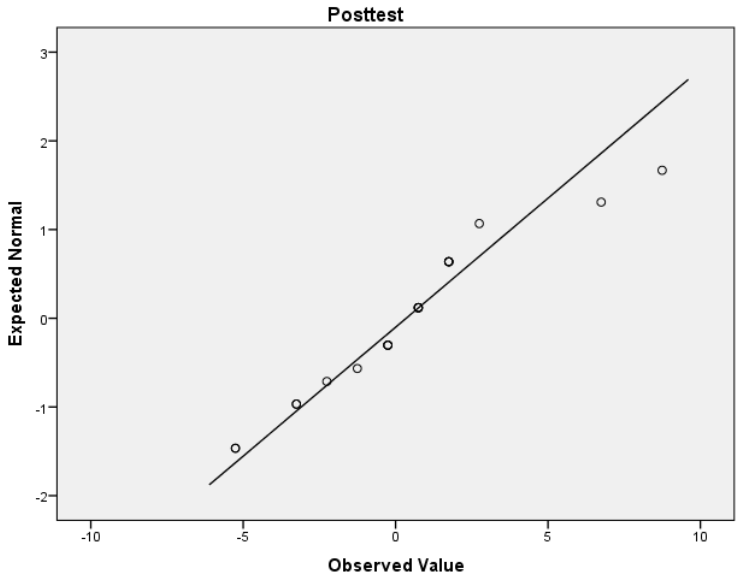
Normal Q-Q Plots for Residuals in Group 1Cog. at Posttest



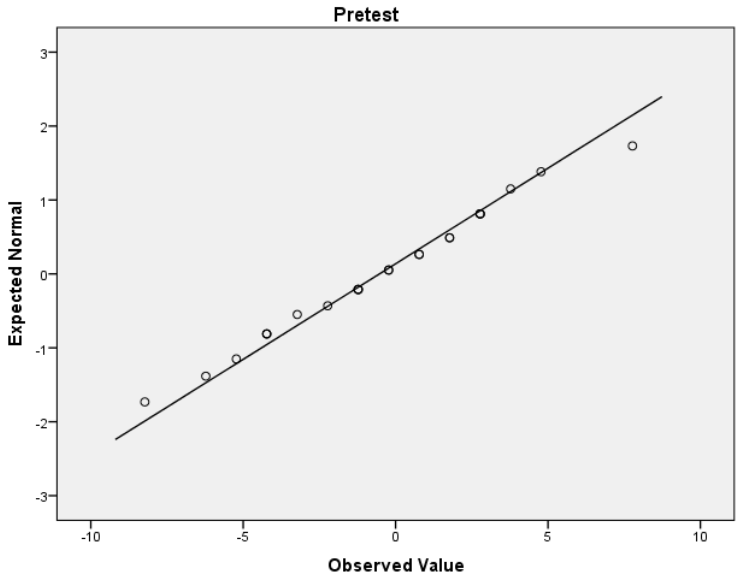
Normal Q-Q Plots for Residuals in Group 2Sit. at Pretest



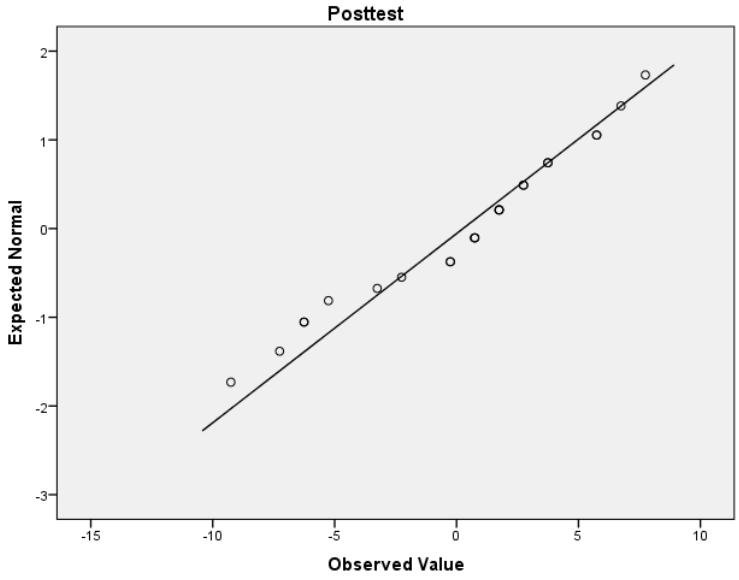
Normal Q-Q Plots for Residuals in Group 2Sit. at Posttest



Normal Q-Q Plots for Residuals in Group 2Cog. at Pretest



Normal Q-Q Plots for Residuals in Group 2Cog. at Posttest



Appendix E: Code Manual

Videoclip 1	
1.1 Nennen Sie die von Ihnen beobachteten Handlungsmaßnahmen (in Stichworten), mit denen die Lehrerin versucht die Aufmerksamkeit der SchülerInnen auf sich zu lenken.	
1F1_K1	**Kriterium "Positionsänderung"**
Code	Handlungsmaßnahme: Änderung der Raumpositionierung, z.B.:
1	- Stellt sich vor die Tafel - Positioniert sich zentral - Nähert sich den SuS
0	Unter Code '1' beschriebene Antwort im Antwortkasten nicht vorhanden
1F1_K2	**Kriterium "Handlungsanweisungen"**
Code	Handlungsmaßnahme: Handlungsanweisung, z.B.:
1	- L. sagt, sie sollen zuhören - "Listen to me" - Sie sagt, was SuS tun sollen - "turn around"
0	Unter Code '1' beschriebene Antwort im Antwortkasten nicht vorhanden
1F1_K3	**Kriterium "Shhh"**
Code	Handlungsmaßnahme: shhh-Laute, z.B.:
1	- shh-Laute - Ruhe-Laute - Zischlaute
0	Unter Code '1' beschriebene Antwort im Antwortkasten nicht vorhanden
1F1_K4	**Kriterium "Gestikulierung"**
Code	Handlungsmaßnahme: Gestikulierung, z.B.:
1	- Gestikulierung - Finger vor den Mund - Hand heben - Winken - Leisezeichen
1F1_K5	**Kriterium "Klatschen"**
Code	Handlungsmaßnahme: Klatschen, z.B.:
1	- klatscht in die Hände - klatscht
0	

1F1_K6	Kriterium "Stille"
Code	Handlungsmaßnahme: Stille, z.B.:
1	- Abwarten - Sie ist still - Unterbricht Sätze - Schweigen
0	

1F1_K7	Kriterium "Direktes Ansprechen"
Code	L. spricht eine S. mit Namen an, z.B.:
1	- direktes Ansprechen - Sie spricht die SuS direkt mit Namen an. - Fatima - L. spricht S. direkt an
0	

1F1_K8	Kriterium "Lauter reden"
Code	Handlungsmaßnahme: Erhobene Stimme, z.B.:
1	- Stimmerhebung - Lauter reden
0	

1F1_K9	Kriterium "Blickkontakt"
Code	Handlungsmaßnahme: Blickkontakt, z.B.:
1	- Schaut in Richtung verschiedener, störender SuS. - Blickkontakt
0	

1F1_K10	Kriterium "Allgemeine Gruppenansagen"
Code	Handlungsmaßnahme: Hinweise, dass L. weitermachen möchte, z.B.:
1	- Kündigt an, dass es weiter geht - L. gibt an, dass es erst weiter geht, wenn es ruhig ist - Weißt darauf hin, dass sie reden möchte - "Boys and girls" - "I'm still waiting" - "Okay"
0	

Ohne Wertung - Beispiele
- L. geht an die Tische - Akustisches Signal (zu vage) - L. berührt S. - L. ruft (zu vage) - L. ermahnt (macht sie nicht) - L. bleibt an der Stelle stehen (zu vage) - Bittet um Aufmerksamkeit - Spricht die ganze Gruppe an - "Be quiet" (sagt sie nicht)) - Verbale Aufforderung (zu vage) - L. ignoriert (SuS ignorieren L. und nicht umgekehrt)

1.3 Welche Probleme versucht die Lehrerin zu vermeiden, indem sie sagt, nur eine/e SchülerIn pro Gruppe soll zur Tafel kommen?

1F3_K1	**Kriterium "Zeitsparen"**
Code	Kommentar: Es geht hierbei darum, dass aus der Antwort abzuleiten istm dass die Lehrkraft versucht zu vermeiden, dass sie Zeit verliert. Problem, Zeitverlust, z.B.:
1	- Zeitverlust - Vermeidet Warten - Schnelle Fortsetzung (des Unterrichts) - Dauert zu lange (grenzwertig) - Geht schneller (grenzwertig)
0	

1F3_K2	**Kriterium "Unruhe / Lärm"**
Code	Kommentar: Es geht hierbei um das allgemeine Problem der Unruhe in der ganzen Klasse. Problem, Unruhe, z.B.:
1	- Sie reduziert die Unruhe. - Unruhe - Durcheinander - Lärm - Chaos
0	

1F3_K3	Kriterium "Platzmangel"
Code	Kommentar: Es geht hierbei darum, dass aus der Antwort abzuleiten ist, dass die Lehrkraft den begrenzten Platz vor der Tafel im Auge hat. Problem, Platzmangel, z.B.:
1	- Gedrängel - Platzmangel - Hindernisse - Stau - Begrenzter Platz
0	

1F3_K4	Kriterium "Verlust Überblick"
Code	Kommentar: Es geht hierbei darum, dass aus der Antwort abzuleiten ist, dass die Lehrkraft nicht mehr gut beurteilen kann, was im Klassenzimmer passiert (Perspektive der Lehrkraft).
1	- Sie vermeidet den Verlust des Überblicks über die Klasse. - Unübersichtlich - Unüberschaubar
0	

Ohne Wertung - Beispiele

- Zu viel Bewegung (zu vage)
- Alle gehen gleichzeitig an die Tafel (Wiederholung der Frage)
- Vermeidung der Ablenkung (zu vage)

1.5 Welche Handlungsmaßnahme der Lehrerin war am erfolgreichsten für die Herstellung der Aufmerksamkeit der ganzen Klasse beim zweiten Phasenübergang?

1F5_K1	Kriterium "direkte Ansprache Schülerin"
Code	Handlungsmaßnahme: Direkte Ansprache Schülerin, z.B.:
1	- direkte Ansprache - "[Name]" - "Turn around" - Ansprache einer S. - Auffordern Einzelner (grenzwertig)
0	

1F5_K2	Kriterium "Abwarten"
Code	Handlungsmaßnahme: Abwarten, z.B.:
1	- Ansage, es geht erst weiter, wenn es ruhig ist - "I'm still waiting" - Abwarten - Abwarten, bis es ruhig ist - Schweigen - Sie ist still
0	

Ohne Wertung - Beispiele

Kodierhinweise für beide Kriterien: Wenn mehrere Handlungsmaßnahmen erwähnt werden, werden nur die ersten zwei, die genannt wurden, kodiert.

- Ermahnen (tut sie nicht)
- Allgemeine Ansprache (zu vage)

1.6 Wie hätte die Lehrerin den Übergang zur Ergebnissicherung an der Tafel so gestalten können, dass sie insgesamt nur einmal hätte warten müssen?

1F6_K1	Kriterium "Instruktion früher"
Code	Kommentar: Hier geht es darum, dass aus der Antwort klar wird, dass die Lehrkraft die Instruktion zu einem früheren Zeitpunkt hätte machen können. Handlungsalternative, frühe Instruktion, z.B:
1	- Instruktion während der Arbeitsphase
0	

1F6_K2	Kriterium "Poster nicht aufhängen"
Code	Kommentar: Hier geht es darum, dass die SuS die Plakate nicht aufhängen mussten.
1	- Die SuS hätten die Poster nicht aufhängen müssen - Präsentation ohne Aufhängen - Präsentation am Sitzplatz
0	

1F6_K3	Kriterium "Plakate abholen"
Code	Kommentar: Hier geht es darum, dass aus der Antwort abgeleitet werden kann, dass die Lehrerin die Plakate abholen konnte. Handlungsalternative: Plakate abholen
1	- Lehrerin holt Plakate ab - Lehrerin sammelt Plakate - Lehrerin hängt Plakate auf
0	

1F6_K4	Kriterium "Präsentation direkt beim Aufhängen"
Code	Kommentar: Hier geht es darum, dass aus der Antwort klar wird, dass die SuS die Ergebnisse direkt nach dem Aufhängen hätten vorstellen können. Handlungsalternative: Präsentation beim Aufhängen, z.B.:
1	- Vorstellen direkt nach dem/ beim Aufhängen - SuS präsentieren direkt an der Tafel - SuS stellen selbst vor (grenzwertig) - SuS können direkt was sagen (grenzwertig)
0	

1F6_K5	Kriterium "Komplette Darstellung Handlungsablauf"
Code	Kommentar: Hier geht es darum, dass aus der Antwort klar wird, dass die Lehrerin das gewünschte SuS-Verhalten komplett hätte darstellen können- Handlungsalternative: Komplette Darstellung Handlungsablauf, z.B.:
1	- Komplette Darstellung / Beschreibung des gewünschten / erwarteten SuS-Verhaltens - "Ich möchte, dass einer aus jeder Gruppe an die Tafel kommt, das Plakat aufhängt und sich dann wieder hinsetzt, währenddessen alle anderen still sind." - All ihre Anweisungen auf einmal stellen - Bekanntgabe der Aufgabestellung (grenzwertig)
0	

Ohne Wertung - Beispiele

Hinweise zum Kodieren aller Kriterien: Bei dieser Frage geht es um Tipps, mit denen die Lehrkraft die Anzahl der Übergangsphasen zurückbringen kann. Tipps, die Sinn machen, aber nicht die Übergangsphasen verringern erhalten keinen Punkt.

- Klarer formulieren
- Nach einander einzeln nach Vorne bitten

1.8 Die Lehrerin gibt an, dass noch wenig Zeit übrig ist. Aus welchen Gründen hält sich die Lehrerin trotzdem an die Regel, erst mit der Instruktion anzufangen, wenn sie die Aufmerksamkeit aller SuS hat?

1F8_K1	Kriterium "Zeitsparen"
Code	Grund: Zeitsparen, z.B.:
1	- Sie spart Zeit - Vermeidet Instruktionswiederholung - Keine Wiederholung
0	

1F8_K2	Kriterium "Gegenseitige Ablenkung"
Code	Grund: SuS lenken sich gegenseitig ab, z.B.:
1	- vermeidet, dass SuS einander ablenken
0	
1F8_K3	**Kriterium "Untergrabung Authorität"**
Code	Grund: Untergrabung L-Authorität, z.B.:
1	- Vermeidet Authoritätsuntergrabung - Auffordern von (gegenseitigem) Respekt
0	
1F8_K4	**Kriterium "Stimme schonen"**
Code	Grund: Schonung der Stimme, z.B.:
1	- Überanstrengung der Stimme - Schonung der Stimme
0	
1F8_K5	**Kriterium "Wahrnehmbare Instruktion"**
Code	Grund: Instruktion von allen SuS warhnehmbar
1	- Instruktion muss wahrnehmbar sein - Alle können die Instruktion mitbekommen - Sicherstelllen, dass alles mitbekommen wird. - Sonst gibt es keinen Lerneffekt - Sonst können die SuS die L. nicht hören
0	
1F8_K6	**Kriterium "Vermeidung Inkonsequenz"**
Code	Grund: Inkonsequenzvermeidung, z.B.:
1	- Sie will Inkonsequenz vermeiden - An der Regel festhalten - SuS müssen nicht denken, dass der Unterricht trotz hoher Lautstärke durchgeht.
0	
Ohne Wertung - Beispiele	
- Weil sonst Unruhe entsteht	

Videoclip 2	
2.1 Welche potentiellen Störungsquellen behebt die Lehrerin, bevor sie mit der ersten Übung anfängt?	
2F1_K1	**Kriterium "Wasserflasche"**
Code	Quelle: Wasserflasche, z.B.:
1	- Flasche - Wasserflasche
0	
2F1_K2	**Kriterium "Wegrücken"**
Code	Quelle: Wegrücken, z.B.:
1	- Sie bittet S. sich vom Nachbarn zu distanzieren - Wegrücken (grenzwertig) - Sie fordert einen Schüler auf vom Mitschüler wegzurücken.
0	
2F1_K3	**Kriterium "Beendung der SuS-Gespräche"**
Code	Quelle: Unruhe in der Klasse, z.B.:
1	- Unruhe / Schwätzen - Sie beruhigt die Klasse / Wartet, bis es ruhig ist. - Sie beendet Schülergespräche - Sie sorgt/ bittet für Ruhe
0	
2.2 In welchen Hinsichten ist es der Lehrerin für den Start der Stunde hilfreich, dass das Datum schon an der Tafel steht?	
2F2_K1	**Kriterium "Nicht umdrehen"**
Code	
1	- Sie muss nicht mit dem Rücken zur Klasse stehen - Sie muss sich nicht umdrehen (grenzwertig) - Sie muss dadurch beim Tafelanschreiben ihren Rücken nicht zur Klasse drehen.
0	

2F2_K2	Kriterium "Aufmerksamkeitsrichtung"
Code	
1	- Um vor Beginn der Übung die Aufmerksamkeit zu bekommen - Sie richtet damit schon vor Anfang des Unterrichts die Aufmerksamkeit der SuS.
0	

2F2_K3	Kriterium "Muss nicht mehr suchen / fragen"
Code	
1	- Sie muss nicht mehr nach dem Datum fragen/ suchen. - Auch wenn SuS das Datum nicht auswendig wissen, können sie direkt mitmachen.
0	

2F2_K4	Kriterium "Zeitersparnis"
Code	
1	- Sie spart Zeit. - Sie kann sofort anfangen. - Sie hat das Unterrichtsmaterial direkt zur Verfügung. - Sie kann inhaltlich direkt an das Datum anknüpfen.
0	

2F2_K5	Kriterium "Unterrichtsroutine"
Code	
1	- Routine - Das ist Teil der Routine mit der der Unterricht immer / öfters anfängt. - Funktion einer Routine: "Eine Routine sorgt für einen fließenden Übergang in die Stunde."
0	

2.4 Mit welchen Handlungsmaßnahmen versucht die Lehrerin das unaufgeforderte Rufen des Datums einiger SuS zu unterbinden?

2F4_K1	Kriterium "Ignorieren"
Code	Handlungsmaßnahme: Ignorieren, z.B.:
1	- Sie ignoriert (gezielt) das Rufen einiger SuS. - Sie schweigt, bis es ruhiger wird.
0	

2F4_K2	Kriterium "Gestikulierung"
Code	Handlungsmaßnahme: Gestikiulierung, z.B.:
1	- Sie gestikuliert - Leisezeichen - Sie legt den Finger auf den Mund.
0	
2F4_K3	**Kriterium "Persönliche Ansprache"**
Code	Handlungsmaßnahme: Direkte Ansprache, z.B.:
1	- Sie spricht S. persönlich an. - "Du bist jetzt nicht dran!" - Das Nennen von Namen - Sie ermahnt SuS
0	
2F4_K4	**Kriterium "Auswahl"**
Code	Handlungsmaßnahme: Auswahl, wer darf, z.B.:
1	- L. wählt S. aus - Sie wält aus, wer das Datum nennen darf. - Sie nimmt SuS dran, die sich melden.
0	
2F4_K5	**Kriterium "Shhh"**
Code	Handlungsmaßnahme: Shhh-Laute, z.B.
1	- Sie macht 'Shh-Laute' - Shhh
0	
Ohne Wertung - Beispiele	
- Warten - Beruhigen (zu vage) - Sie bittet um Stille (zu vage) - Passagen, die sich auf die Situation "Yessin" beziehen.	

2.6 Die Lehrerin fragt einen Schüler insgesamt zweimal nach dem Datum. Was will die Lehrerin dem Schüler deutlich machen, indem sie ihn das zweite Mal dran nimmt?	
2F6_K1	**Kriterium "Besser zuhören"**
Code	Message an S.: Er soll seinem Klassenkameraden besser zuhören, z.B.:
1	- Er soll auch aufmerksam sein, wenn er nicht dran ist. - Er soll besser zuhören. - Er soll besser aufpassen. - Du bist noch im Fokus meiner Aufmerksamkeit. - Kontrolle, ob er zugehört hat. - L. nimmt ihn extra dran, weil er nicht aufgepasst hat.
0	
2F6_K2	**Kriterium "Demonstration der Klasse"**
Code	Message an Klasse: Ich sehe es, wenn jemand nicht aufpasst, z.B.:
1	- Monitoring / Withitness - Ich sehe alles! - Klasse / SuS sollen aufpassen - Ich sehe es, wenn jemand nicht aufpasst.
0	
2F6_K3	**Kriterium "Regeln"**
Code	Kommentar: Nur einen Punkt vergeben, wenn Hinweis auf generelle Prinzipien / Regeln gemacht wurde. Message an Klasse: Es gelten Regeln, z.B.:
1	- Es gelten Regeln an die man sich halten muss.
0	
2F6_K4	**Kriterium "Interesse an Lernerfolg"**
Code	Message an S.: L. möchte, dass S. etwas lernt, z.B.:
1	- Ich gebe dir eine weitere Chance - Du lernst nichts, wenn du nicht zuhörst. - Du sollst das richtige Ergebnis lernen.
0	

2.7 Woran kann man erkennen, dass es den SuS erlaubt ist während der Stunde zu trinken?	
2F7_K1	**Kriterium "Trinkbecher"**
Code	
1	- Es stehen Trinkbecher auf den Tischen. - Getränke auf dem Tisch - Gläser / Becher
0	
2F7_K2	**Kriterium "Wasserflasche"**
Code	
1	- Am Anfang der Stunde stand eine Wasserflasche auf dem Tisch.
0	
Ohne Wertung - Beispiele	
- Dass eine Wasserflasche entfernt wurde	
Videoclip 3	
3.2 Aus welchen Gründen gibt die Lehrerin sich Mühe, die Ereignisse im Türbereich schnell zu beenden?	
3F2_K1	**Kriterium "Zeitsparen"**
Code	
1	- So kann sie die Stunde / das Begrüßungsritual/ der Unterricht schneller beginnen. - Sie möchte pünktlich beginnen - Der Unterricht verzögert sich sonst - So verliert sie wenig Zeit.
0	
3F2_K2	**Kriterium "Keine Störungen"**
Code	
1	- So gibt sie den SuS keine Chance, Störungen zu entwickeln. - Ablenkungen sollen vermieden werden (grenzwertig) - Klasse wird lauter - Um Störungen zu vermeiden
0	

3F2_K3	Kriterium "Kein Zutritt"
Code	
1	- Sie will unbefugten SuS den Zutritt zum Klassenzimmer unmöglich machen. - Um Störungen von außen zu vermeiden (grenzwertig)
0	

Ohne Wertung - Beispiele
- Damit sich anfangen kann (zu vage) - Versucht Streit / Diskussion zu verhindern - schnell (Grund fehlt) - Pünktlichkeit (Grund fehlt)

3.3 Aus welchen Aussagen der Lehrerin kann man schließen, dass die SuS feste Sitzplätze haben?

3F3_K1	Kriterium "Dein Platz"
Code	
1	- Sie macht eine persönliche Ansage ("Auf deinen Platz, Dennis."). - Sie bittet die SuS auf ihre (jeweiligen) Plätze zu gehen.
0	

3F3_K2	Kriterium "Verweis auf geänderten Sitzplan"
Code	Kommentar: Nur angaben, die auf Änderungen im Sitzplan verweisen, z.B.:
1	- "Habt ihr das schon wieder geändert?" / "So blicke ich gar nicht mehr durch, wer wo sitzt." - Habt ihr euch schon wieder umgesetzt (grenzwertig) - Sie erwähnt den geänderten Sitzplan
0	

Ohne Wertung - Beispiele
- Sie erwähnt den Sitzplan / Die Sitzordnung (zu vage)

3.5 Welche Gründe könnte die Lehrerin dafür haben, dass sie die SuS auffordert aufzustehen, bevor sie mit dem Grußaustausch anfängt?

3F5_K1	Kriterium "Schnelle Übersicht"
Code	
1	- So bekommt sie eine schnelle Übersicht.
0	

3F5_K2	Kriterium "Körperliche Aktivierung"
Code	
1	- Durch die körperliche Bewegung aktiviert sie die SuS.
	- Das Aufstehen aktiviert
	- Anregung des Kreislaufs / Kreislauf wird angeregt
0	

3F5_K3	Kriterium "Gleichheit der SuS"
Code	
1	- Durch die homogenisierte Körperhaltung zeigt sie die Gleichheit aller SuS. - Für das Gemeinschaftsgefühl
0	

3F5_K4	Kriterium "Disziplin / Höflichkeit / Respekt"
Code	
1	- So fordert sie die Disziplin / Höflichkeit der SuS ein. - Um Authorität zu zeigen
0	

3F5_K5	Kriterium "Stundenanfang"
Code	
1	- So zeigt sie, dass sie anfangen möchte. - Jeder kann sich auf den (Stunden-)Anfang einstellen - Signalisierung des Stundenanfangs - Symbolischer Beginn der Stunde
0	

3F5_K6	Kriterium "Regeln gelten für alle"
Code	Kodierhinweise, nur eine 1, wenn Bedingung erfüllt.
1	- So zeigt sie, dass die Unterrichtsregeln für alle SuS gleichermaßen gelten.
0	

Ohne Wertung - Beispiele

- Sie will dass di SuS aufmerksam sind (zu vage)
- Dass die L. es ernst meint
- Weil es Teil des Begrüßungsrituals ist
- Das ist eine Routine / Ritual
- Gemeinsamer Start
- Ritual sorgt für Ruhe
- Ruhe
- Die Regeln gelten für alle, nicht nur für L.

3.6 Beschreiben Sie in Stichworten die verschiedenen Phasen des Grußaustausches, mit dem die Stunde anfängt.	
3F6_K1	**Kriterium "SuS stehen"**
Code	Kodierhinweise, nur eine 1, wenn Bedingung erfüllt.
1	- Die SuS stehen auf / stehen an ihren Plätzen.
0	
3F6_K2	**Kriterium "Begrüßung"**
Code	Kommentar: Nur ein Punkt, wenn gegenseites Grüßen erwähnt, z.B.:
1	- Die Lehrerin begrüßt die SuS / sagt "Good morning", SuS erwidern den Gruß - Gegenseitige Begrüßung - L. und SuS begrüßen einander
0	
3F6_K3	**Kriterium "Hinsetzen und Lied"**
Code	Kommentar: Nur ein Punkt wenn hinsetzen und Lied / singen erwähnt
1	- Die SuS setzen sich hin und singen ein (begrüßungs-)Lied.
0	
3F6_K4	**Kriterium "Wie gehts"**
Code	
1	- L. sagt, ihr gehe es gut und fragt, wie es SuS geht - L. sagt, ihr gehe es gut - Die Lehrerin fragt, wie es den SuS geht. - L. erwidert Frage (wie es geht) - SuS sagen, ihnen gehe es auch gut
0	
Ohne Wertung - Beispiele	
- Begrüßung (zu vage; Wer begrüßt wen?) - "Good Morning" sagen (zu vage)	
3.7 Benennen Sie in Stichworten die Wirkungen des Grußaustausches auf das Verhalten der SuS.	
3F7_K1	**Kriterium "SuS sitzen"**
Code	
1	- Die SuS sitzen alle (an ihrem Platz). - SuS befinden sich an ihren Plätzen.
0	

3F7_K2	Kriterium "SuS sind ruhig"
Code	
1	- SuS haben private Unterhaltungen eingestellt. - Die SuS werden leise / kommen zur Ruhe.
0	

3F7_K3	Kriterium "L. hat Aufmerksamkeit"
Code	
1	- Die SuS schauen alle auf die Lehrerin. / Die Lehrerin hat die Aufmerksamkeit. - SuS sind im Unterrichtsgeschehen angekommen (grenzwertig) - Den SuS ist es klar, dass die Stunde anfängt - SuS sind auf Aufgabe konzentriert - Niemand kann sich entziehen - SuS sind auf Unterrichtssituation eingestimmt - SuS sind in der Stunde angekommen
0	

3F7_K4	Kriterium "Körperliche Aktivierung"
Code	
1	Die SuS werden körperlich aktiviert.
0	

Ohne Wertung - Beispiele
- Sie halten sich an Regeln.

3.8 Nennen sie die Vorteile der Tischordnung hinsichtlich der Prävention von Unterrichtsstörungen durch die SuS

3F8_K1	Kriterium "L. kann alle S. sehen"
Code	Kodierhinweise, nur eine 1, wenn Bedingung erfüllt.
1	- Die SuS sind für die Lehrerin gut wahrnehmbar.
0	

3F8_K2	Kriterium "SuS können L. und / oder Tafel sehen"
Code	
1	- SuS sind frontal ausgerichtet - Blickrichtung nach Vorne - Niemand sitzt mit Rücken zur L. - Die SuS können die Lehrerin gut wahrnehmen. - Die SuS können die Tafel gut wahrnehmen. - Keiner muss sich drehen um die Lehrkraft / Tafel zu sehen
0	

3F8_K3	Kriterium "Schwierige Unterhaltung"
Code	Kommentar: Nur einen Punkt vergeben, wenn Vorteil von Tischordnung genannt
1	- SuS können relativ schwer ins Gespräch kommen. - SuS können nur mit Sitznachbar reden - SuS können leichter alleine platziert werden - Es ist für SuS schwieriger miteinander zu reden
0	

3F8_K4	Kriterium "Gänge frei"
Code	Kommentar: Nur einen Punkt vergeben, wenn Vorteil der freien Gänge genannt.
1	- Die Gänge sind frei, es gibt keine Gefahr des Stolperns / Drängelns. - L. kann sich einfacher durch die Klasse bewegen. - SuS finden schneller ihre Plätze (grenzwertig)
0	

Videoclip 4

4.1 Beschreiben Sie in Stichworten das Verhalten, mit dem der Lehrer eine neue Unterrichtsphase einleitet.

4F1_K1	Kriterium "Ausruf"
Code	Kodierhinweise, nur eine 1, wenn Bedingung erfüllt.
1	- Er sagt: "So..." - Ausruf - "So"
0	

4F1_K2	Kriterium "Positionierung"
Code	
1	Er positioniert sich zentral vor der Tafel.
0	

4F1_K3	Kriterium "Gestik"
Code	
1	- Er hebt die Hand / den Arm. - Handzeichen (grenzwertig)
0	

4F1_K4	Kriterium "Runterzählen"
Code	Kodierhinweise, nur eine 1, wenn Bedingung erfüllt.
1	- Er sagt: "3,2,1, stop". / Er zählt runter. - Countdown
0	

Ohne Wertung - Beispiele	
- Er zählt bis 3 - Er benutzt Gestik (zu vage) - Zischlaute	
4.2 Nennen Sie in Stichworten die Handlungsmaßnahmen, mit denen der Lehrer die SuS zum Flüstern bewegt.	
4F2_K1	**Kriterium "Routine"**
Code	
1	- Er verweist auf eine Routine. - "Wie wir oft gemacht haben."
0	
4F2_K2	**Kriterium "Flüstert selber"**
Code	
1	- Er macht das Flüstern vor - Er fängt selbst an zu flüstern. - Er spricht leiser. - Vormachen - Senken der Stimme
0	
4F2_K3	**Kriterium "Whisper"**
Code	
1	- Er sagt flüstern auf Englisch - Er sagt (mehrmals) "whisper".
0	
4F2_K4	**Kriterium "Gestik"**
Code	
1	- Er hebt beide Hände. - Senken der Arme - Er macht beruhigende Gestiken. - Er macht eine passende Gestik (grenzwertig) - Gestik (grenzwertig) - Handbewegung (grenzwertig)
0	

4F2_K5	Kriterium "Shh"
Code	
1	- Er macht "Shh"-Laute. - "Shh"
0	
4F2_K6	**Kriterium "Beschreibt SuS-Verhalten"**
Code	
1	- Er beschreibt das gewünschte SuS-Verhalten - Verbale Aufforderung - "Wir lesen leise mit" - Er sagt, sie sollen flüstern.
0	
Ohne Wertung - Beispiele	
- Es scheint öfters zu passieren. (Keine Handlungsmaßnahme)	
4.3 Woran kann man erkennen, dass die Übung zum Textverstehen öfters in der gesehenen Form stattfindet?	
4F3_K1	**Kriterium "Wie wir oft gemacht haben"**
Code	
1	- Der Lehrer sagt: "Wie wir schon oft gemacht haben". - Der Lehrer äußert dies
0	
4F3_K2	**Kriterium "Minimale Instruktionen"**
Code	Faktor geringe Instruktion
1	- Der Lehrer gibt minimale Instruktionen.
0	
4F3_K3	**Kriterium "Kein Nachfragen"**
Code	Faktor Zeit, geringe Anzahl Nachfragen
1	- Die SuS machen (fast) alle sofort mit. - SuS steigen sofort drarauf ein - Es gibt keine Nachfragen. - Manche SuS nehmen schon bei seiner Arbeitsanweisung das Buch in die Hand.
0	

Ohne Wertung - Beispiele
- Die SuS sind damit vertraut - Die SuS wissen, was zu tun ist (woran erkannt man das?) - Dass alle SuS mitarbeiten (zu vage) - Es scheint öfters zu passieren. (Keine Handlungsmaßnahme)
4.4 Welche Vorteile hat die gesehene Arbeitsform hinsichtlich der Vermeidung potentieller Unterrichtsstörungen durch die SuS?

4F4_K1	**Kriterium "Schauen in Bücher"**
Code	Nur 1, wenn Blickwinkel (Kopf) und Interaktion mit Mitschülern genannt
1	- SuS schauen ins Buch und lenken einander ab.
0	

4F4_K2	**Kriterium "L. kann SuS beobachten"**
Code	
1	- L. kann SuS beobachten - L. kann SuS während der Übung gut beobachten - Es fällt dem Lehrer schneller auf, wenn SuS nicht mitmachen.
0	

4F4_K3	**Kriterium "Leise durch Tonband"**
Code	
1	- Das Tonband stellt einen Reiz dar, keine anderen Geräusche zu machen. - Durch Tonband Ruhe - Um CD hören zu können, müssen die SuS selber still sein.
0	

4F4_K4	**Kriterium "Fällt auf, wenn keine Beteiligung"**
Code	
1	- Es fällt auf, wenn sich die SuS nicht arbeiten. - Off-task / stören fällt auf
0	

4F4_K5	**Kriterium "Volle Beanspruchung"**
Code	
1	- Alle Sinne beansprucht - Fokusiert durch mitlesen - Beschäftigung durch flüstern und hören - Es bleibt nicht viel Gelegenheit für eine unterrichtsferne Beschäftigung. - Die SuS müssen sich sehr konzentrieren und können parallel keine anderen S. ansprechen
0	

Ohne Wertung - Beispiele
- Jeder hat etwas zu tun - Jeder ist beschäftigt / konzentriert - Jeder schaut ins Buch

4.6 Welche Gründe könnte der Lehrer dafür aben, erst zum Tisch der Schülerin zu gehen, bevor er sie tadelt?

4F6_K1	**Kriterium "Keine Ablenkung anderer SuS"**
Code	
1	- Er vermeidet durch lautes Reden andere SuS abzulenken. - Er vermeidet, das andere SuS ihre Aufmerksamkeit auf S. richten.
0	

4F6_K2	**Kriterium "Geringe Bloßstellung"**
Code	
1	- Die Schülerin wird weniger bloßgestellt. - Geringere Bloßstellung
0	

4F6_K3	**Kriterium "Bessere Beurteilung"**
Code	
1	- So kann er die Situation genauer beurteilen. - Er weiß selber nicht, was die S. macht. - Um nachzuschauen, was sie macht.
0	

4F6_K4	**Kriterium "Dominanz"**
Code	
1	- Seine körperliche Nähe erzeugt Dominanz. - Effekt ist größer wenn er näher an S. ist (grenzwertig)
0	

Ohne Wertung - Beispiele
- Nicht öffentlich (warum ist das besser?) - Er muss nicht laut werden (Warum ist das besser?) - Körperliche Präsenz (Warum ist das besser?) - Um nicht laut reden zu müssen (Warum ist das besser?)

4.7 Was stellt sich im Zwiegespräch als Hauptgrund dafür heraus, dass die Schülerin sich zunächst nicht an der Übung beteiligte?	
4F7_K1	**Kriterium "Andere Materialien"**
Code	
1	- Die S. hatte andere Materialien vor sich / in der Hand. - Sie hat etwas Anderes gelesen.
0	
Ohne Wertung - Beispiele	
- Unterstreichen - Die S. hat etwas andere gemacht (zu vage) - Die S. hat etwas andere gemacht (zu vage)	

DEUTSCHE ZUSAMMENFASSUNG

Einführung

Obwohl die große Bedeutung von gut ausgebildeten Lehrkräften allgemein anerkannt ist, zeigt die Forschung, dass die Wirksamkeit von Lehramtsstudiengängen in dieser Hinsicht nicht immer selbstverständlich ist (Darling-Hammond, 1999; Kennedy et al., 2008; Rauin, 2011). Der Übergang von der universitären ersten Phase der Ausbildung zur Unterrichtspraxis wird von den angehenden Lehrkräfte oft als ein „Praxisschock" erfahren (Müller-Fohrbrodt et al., 1978; Veenman, 1984). Während Wissen im Rahmen des Studiums, zum Beispiel in Klausuren oder Hausarbeiten, von Studierenden überwiegend mühelos theoretisch wiedergegeben werden kann, ist ein Transfer dieses universitären Wissens in die Schulpraxis nicht selbstverständlich (Korthagen & Kessels, 1999; Mägdefrau & Schumacher, 2001; Renkl, 1996). Diese Tatsache stellt die universitäre Lehrkräftebildung vor die generelle Frage nach den Lernmechanismen von Lehramtsstudierenden einerseits. Andererseits gilt es herauszufinden, wie effektive Verbindungen zwischen Praxiserfahrungen und universitären Lerngelegenheiten hergestellt werden können (Zeichner, 2009).

Illustrativ für die Herausforderung, universitäre Lerngelegenheiten so zu konzipieren, dass sie auch auf die Unterrichtspraxis vorbereiten, ist der Unterrichtsaspekt der Klassenführung (Havers, 2010). Die Forschung in Bezug auf das Konzept der Klassenführung, im Englischen *Classroom Management*, hat seine Wurzeln in der pädagogischen Forschung der USA ab Beginn der zweiten Hälfte des 20sten Jahrhunderts (Brophy, 2006; Haag & Streber, 2012). Im *Handbook of Classroom Management* definieren Evertson und Weinstein den Begriff Klassenführung als „the actions teachers take to create an environment that supports and facilitates both academic and social-emotional learning" (Evertson & Weinstein, 2006b, p. 4). Aus dieser Definition lassen sich drei wichtige Aspekte zur Klassenführung ableiten: Erstens dass es sich bei Klassenführung in erster Linie um das sichtbare in-situ Verhalten der Lehrkraft handelt und nicht um anderweitige unterrichtsbezogene Fähigkeiten, wie zum Beispiel die Fähigkeit Lernmaterialien zu gestalten, Unterrichtseinheiten zu planen, Schülerleistungen zu beurteilen oder einzelne Schüler zu beraten. Zweitens zeigt die Definition, dass Klassenführung sich auf das Schaffen von Lernbedingungen konzentriert und nicht auf die Lernaktivitäten an sich. Aus diesem Punkt lässt sich schließen, dass Klassenführung unabhängig vom Unterrichtsfach ist. Drittens folgt aus der Definition, dass Klassenführung

lernunterstützend ist und nicht ein Lernziel an sich darstellt, wie zum Beispiel das Beibringen von Disziplin. Die Forschung (Doyle, 2006; Evertson & Weinstein, 2006a; Gettinger & Kohler, 2006; Landrum & Kauffman, 2006) ergab, dass sich Klassenführung aus folgenden drei Dimensionen zusammensetzt: Die erste Dimension bezieht sich auf das Managen von Gruppeninteraktion und ist dadurch eng verknüpft mit den Aspekten *Monitoring* (Ophardt & Thiel, 2013), *Withitness* und *Overlapping* (Kounin, 1970) sowie der physischen Einrichtung des Klassenzimmers (Borich, 2011). Die zweite Dimension bezieht sich auf die Maximierung und effektive Verwendung der verfügbaren Lernzeit. Sie bezieht damit Aspekte wie den der konsequenten Einhaltung von Regeln (Emmer et al., 1980) und jenen der effektiven Anwendung von Unterrichtsroutinen (Gettinger & Kohler, 2006) ein. Die dritte Dimension betrifft das Management von Unterrichtsstörungen. In erster Linie geht es hierbei um die Verwendung des sogenannten Low-Profile Ansatzes. Dieser zielt darauf ab, dass die Interventionen einer Lehrkraft auf Unterrichtsstörungen so wenig wie möglich den Unterrichtsfluss unterbrechen (Borich, 2011).

Verschiedene Metastudien haben die Bedeutung der Klassenführungsfähigkeit einer Lehrkraft für die Schülerinnen- und Schülerleistung nachgewiesen (Seidel & Shavelson, 2007; Wang et al., 1993). So wurde empirisch belegt, dass im Unterricht von Lehrkräften, welche erfolgreich Ordnung im Klassenzimmer herstellen und Störungen im Unterricht verhindern können, sich die verfügbare Lernzeit (Helmke, 2007) sowie die Beschäftigung der Schülerinnen und Schüler mit dem Lerngegenstand erhöht (Emmer & Stough, 2001). Diese Effekte erhöhen dann wiederum die Wahrscheinlichkeit, dass erfolgreiches Lernen stattfindet. Dieser Wirkungskette entsprechend wird effektiver Klassenführung eine zentrale Bedeutung in Modellen von gutem Unterricht zugeschrieben (Helmke, 2012; Ophardt & Thiel, 2008). Nicht nur die Schülerinnen und Schüler, sondern auch die Lehrkräfte haben ein klares Interesse daran, Unterrichtsstörungen effektiv vorbeugen zu können. Verschiedene Studien zeigen, dass ein hohes Maß an Unterrichtsstörungen mit niedrigen Werten an lehrkraftseitiger Zufriedenheit und Selbsteingeschätzte Kompetenz sowie mit häufigerem Vorkommen von Burn-Out Fällen einhergehen (Brouwers & Tomic, 2000; Friedman, 2006; König & Rothland, 2016). Vor allem zeigen sich Herausforderungen der Klassenführung für beginnende Lehrkräfte (Veenman, 1984). Viele dieser Lehrkräfte geben an, sich hinsichtlich

dieses Aspekts der Lehrtätigkeit nicht ausreichend genug vorbereitet zu fühlen (Merrett & Wheldall, 1993).

Modellierung von Klassenführungskompetenz

Für die Modellierung von Lehrkräftekompetenz im Allgemeinen wird oft auf das Modell der professionellen Kompetenz zurückgegriffen (Baumert & Kunter, 2011). Da aber die Bedeutung von deklarativem Wissen bei Klassenführung eine geringe Rolle zu spielt (Cocard & Krähenbühl, 2015), ist das Modell der professionellen Kompetenz für die Modellierung von Klassenführungskompetenz im Speziellen weniger geeignet. Rezente Initiativen binden den Kontext des Unterrichts stärker in die Kompetenzmodellierung ein (König, 2015a). Dabei wäre vor allem auf das PID-Model hinzuweisen (Blömeke et al., 2015b). Dieses Model versteht Kompetenz als ein Kontinuum, in dem Unterrichtsperformanz einerseits und mentale Dispositionen wie Wissen und Überzeugungen andererseits die beiden Enden bilden. Zwischen diesen beiden Enden, also aufbauend auf den mentalen Dispositionen und prädiktiv für Unterrichtsperformanz, befinden sich die drei situationsspezifischen Fähigkeiten *Perception* (Wahrnehmung)*, Interpretation* (Interpretation), und *Decision-Making* (Entscheidungsfähigkeit). Das PID-Modell versteht sich als ein übergreifendes Konzept, das andere ebenfalls situierte Kompetenzmodelle wie beispielsweise das der *professionellen Unterrichtswahrnehmung* (Stürmer & Seidel, 2017), das *Noticing* (Sherin et al., 2011a) und die *Analysekompetenz* (Biaggi, Krammer, & Hugener, 2013) umfasst (Stahnke et al., 2016). Diese situationsspezifischen Fähigkeiten können durch die Analyse von Videovignetten zum Ausdruck gebracht werden. Für die Erfassung solcher situationsspezifischer Fähigkeiten werden deswegen vor allem Videovignettentests eingesetzt (Lindmeier, 2013; Seidel & Thiel, 2017, 2017).

Die Vermittlung von Klassenführungskompetenz an der Universität

In den letzten Jahren hat das Thema Klassenführung in vielen deutschen Universitäten an Bedeutung gewonnen (Lohmann et al., 2011). Dem Thema wird nun von einer Mehrheit der Dozentinnen und Dozenten in der Lehrerbildung eine bedeutende Rolle zugeschrieben (Kunina-Habenicht et al., 2012). Typische Initiativen, Klassenführungskompetenz von Lehramtsstudierenden im Rahmen ihrer universitären Ausbildung zu stärken, umfassen vor allem die Begleitung von Praxisphasen sowie die Analyse von Videovignetten (Stough, 2006).

Praxisphasen wird in der ersten Phase der Lehrerbildung von allen Beteiligten – Dozentinnen und Dozenten, Studierenden und Politikern – eine sehr hohe Bedeutung zugeschrieben (Holtz, 2014). Auch für die Vermittlung von Klassenführungskompetenz wird oft auf die Reflexion von Praxiserfahrung gesetzt (Larrivee, 2006). Allerdings zeigt sich, dass ein Praktikum keine einheitliche und konsistente Lerngelegenheit für Lehramtsstudierende darstellt (Gröschner et al., 2015). Zusätzlich ist die empirische Forschung zur Wirksamkeit von Praxisphasen oft fragmentarisch und basiert sich oft auf durch Beteiligte subjektiv selbsteingeschätzte Kompetenz (Hascher, 2012a). Dennoch gibt es eine Vielfalt an empirischen Studien. Sie zeigen unter anderem dass Lehramtsstudierende während ihr Praktikum eine realistischere Vorstellung vom Beruf der Lehrkraft entwickelnd (Hascher, 2012b). Zusätzlich führt die Teilnahme an Schulpraktika im Durchschnitt zu einer Steigerung professioneller Unterrichtswahrnehmung (Mertens & Gräsel, 2017) und pädagogischen Wissen (König, 2013). Darüber hinaus zeigt die jüngere Forschung, dass Lehramtsstudierende oft nebenberuflich erhebliche zusätzliche Praxiserfahrungen sammeln (Bäuerlein & Reintjes, 2018), was die oben beschriebene Effekte verstärken würde.

Das Thema Klassenführung wird in Universitäten oft über die Analyse von Videovignetten behandelt (Jones, 2006). Videovignetten dienen dazu, Studierenden Einblicke in realistischen Unterricht zu geben (Krammer & Reusser, 2004). Sie ermöglichen es ihnen, theoriebasiert zu reflektieren und zu analysieren, während gleichzeitig die Unmittelbarkeit einer authentischen Unterrichtssituation ausgeklammert wird (Helmke & Helmke, 2004). Somit werden Unterrichtsvideos eine wichtige Rolle für die Verknüpfung von Theorie und Praxis zugeschrieben (Blomberg et al., 2013; Krammer, 2014; Santagata & Guarino, 2011). Verschiedene Studien haben bereits mehrmals die Wirkung der Analyse von Videovignetten in der ersten Phase der Lehrkräfteausbildung nachgewiesen (Gaudin & Chaliès, 2015; Seidel & Thiel, 2017; Steffensky & Kleinknecht, 2016). Darüber hinaus haben andere Studien deren Wirksamkeit für die spezifische Vermittlung von Klasseführungskompetenz bestätigen können (Gold et al., 2017; Kramer et al., 2017).

Eine wichtige Frage, die in den letzten Jahren in den Fokus der Forschung zur Wirksamkeit von Lehrerbildung gerückt ist, bezieht sich auf die didaktische Einbettung von Videovignetten (Blomberg et al., 2013). Sollte ein Video eher als Anker dienen, der die vorher gesammelten Erfahrungen, die subjektiven Theorien und das implizites Wissen von Lehramtsstudierenden

aktiviert (Neuweg, 2004) und es ihnen damit ermöglicht, Regelmäßigkeiten, Strategien und Theorien induktiv abzuleiten (Blomberg et al., 2013)? Oder sollten Videos eher verwendet werden, um vorher vermitteltes Wissen über stark strukturierte Analyseschritten deduktiv anzuwenden (Goeze et al., 2013)? Eng mit dieser Frage verbunden ist die Diskussion über die Menge an kognitiver Belastung (*cognitive load)* (Sweller, 1988), die mit dem Lernen mit Videovignetten assoziiert wird (Carter et al., 1988; Sabers et al., 1991). Sollte einem kognitivistischen didaktischen Ansatz folgend die Komplexität der zu lernenden Materie über eine stark strukturierte Lerngelegenheit deutlich verringert werden, um damit kognitive Überlastung zu vermeiden? Oder führt eine solche methodische und theoretische Strukturierung nur zum Erwerb von sogenanntem „trägen Wissen", das sich nicht in die Praxis umsetzen lässt (Renkl, 1996)? Daraus würde folgen, dass videobasierte Seminare alternativ eher im Sinne eines situierten didaktischen Ansatzes, der vor allem auf die Aktivierung, Reflexion und Diskussion von eigenen Erfahrungen und subjektiven Theorien abzielt, eingesetzt werden sollten (Korthagen, 2010).

Erste Ergebnisse der bislang wenigen Studien, die die Effekte beider didaktischer Ansätze für die Einbettung von Videovignetten systematisch vergleichen, zeigen, dass beide Ansätze effektiv sind, aber unterschiedliche Auswirkungen haben. So scheint für die Vermittlung von deklarativem Wissen der kognitivistische Ansatz besser geeignet zu sein (Kumschick et al., 2017; Seidel et al., 2013). Für die Vermittlung von situierten Kompetenzaspekten, wie dem Erstellen von Unterrichtsentwürfen und der Förderung der Selbstreflexion dagegen, wird ein situierter Ansatz als besser geeignet ausgewiesen (Blomberg et al., 2014; Seidel et al., 2013). Konträr zu den Erwartungen aus der Theorie wurden in diesen Studien keine Unterschiede hinsichtlich der kognitiven Belastung (Syring et al., 2015), der selbsteingeschätzten Klassenführungskompetenz (Kumschick et al., 2017) oder der Motivation der Lernenden (Syring et al., 2015) gefunden. Allerdings gaben Lehramtsstudierende eines Videoseminars mit situiertem Ansatz höhere Werte der Lernvertiefung (*Immersion*) an als Studierende, die ein auf Basis des kognitivistischen Ansatzes aufgebautes Seminar besuchten (Syring et al., 2015). Die Erkenntnisse dieser Studien zeigen, dass beide Ansätze in der universitären Lehrerbildung ihre Berechtigung haben.

Forschungsdesiderat

Während allgemein davon ausgegangen wird, dass Praxiserfahrungen von Lehramtsstudierenden für die Vermittlung von Klassenführungskompetenz eine wichtige Rolle spielen, ist über die Art und Weise, wie videobasierte Seminare das Vorhanden- oder Nichtvorhandensein solcher Erfahrung thematisieren können, wenig bekannt. Aus der Theorie zu situiertem Lernen und kognitiver Belastung lässt sich die Vermutung ableiten, dass die Verfügbarkeit von Praxiserfahrungen einen erheblichen Einfluss auf den Erwerb von Klassenführungskompetenz haben könnte. Denn nur von einer Person, die eine realistische Vorstellung von Schule und Unterricht hat, kann erwartet werden, dass sie im Rahmen eines Lehrangebotes ohne strukturiertes Vorgehen Regelmäßigkeiten induktiv aus einer Videovignette ableiten kann. Gleichzeitig wäre umgekehrt zu erwarten, dass Lehramtsstudierende ohne Praxiserfahrungen deutlich schneller von der Komplexität von Videovignetten überlastet sind und nicht auf verfügbares Vorwissen zurückgreifen können. Diesen Studierenden müssten demnach im Rahmen einer strukturierten Methode zunächst Theorien vermittelt werden, um sie im zweiten Schritt eine Theorie deduktiv auf die Videovignette anwenden zu lassen.

Ziel und Frage der Studie

Das Ziel dieser Studie liegt darin zu untersuchen, auf welche Weise videobasierte Seminare zur Vermittlung von Klassenführungskompetenz auf die Praxiserfahrungen von Lehramtsstudierenden aufbauen können. Zu dieser Zielsetzung wurde folgende Forschungsfrage formuliert:

Wie kann Klassenführungskompetenz unter Berücksichtigung von Praxiserfahrungen von Lehramtsstudierenden mittels Videovignetten effektiv vermittelt werden?

Aus der Theorie zur Verwendung von Videovignetten in der Lehrerbildung wurden zwei Hypothesen formuliert. 1. Lehramtsstudierende mit Praxissemestererfahrung profitieren stärker von einem situierten didaktischen Ansatz. 2. Konträr dazu profitieren Lehramtsstudierende ohne Praxiserfahrung stärker von einem kognitivistischen didaktischen Ansatz.

Methodologie

Um diese Hypothesen zu prüfen wurde ein quasi-experimentelles Forschungsdesign entworfen, in dem die Lerneffekte von vier videobasierten Seminaren zum Thema

Klassenführung in einem Pretest - Posttest Verfahren erfasst wurden. In der Studie wurden die Faktoren „Praxiserfahrung“ und „Seminartyp“ systematisch kontrolliert. Lehramtsstudierende gehörten einer von zwei Kohorten an. Die erste Kohorte bestand aus Studierenden, die das Seminar direkt nach Abschluss des Praxissemesters besuchten. Die zweite Kohorte bestand aus Studierenden, die noch kein Praxissemester besucht hatten und direkt nach dem Seminar mit der Teilnahme am Praxissemester beginnen würden. Jede Kohorte wurde jeweils randomisiert zwei Seminartypen zugeteilt. Ein Seminartyp basierte auf den didaktischen Prinzipien eines kognitivistischen Ansatzes. Der andere Seminartyp basierte hingegen auf den didaktischen Prinzipien eines situierten Ansatzes.

Die Stichprobe bestand aus insgesamt 88 Studierenden des gymnasialen Studiengangs: 44 der ersten Kohorte verfügten über Erfahrung im Rahmen des Praxissemesters und 44 aus der zweiten Kohorte wiesen keine Praxissemestererfahrung auf. Studierende in der ersten Kohorte waren zu 56% weiblich und durchschnittlich 21.9 (*SD* = 2,3) Jahre alt. Studierende in der zweiten Kohorte waren zu 63,3% weiblich und durchschnittlich 22,1 (SD = 4,5) Jahre alt. Die Kohorten unterschieden sich deutlich in der Anzahl an Monaten, in denen Studierende bereits außeruniversitäre Praxiserfahrungen z.B. als Vertretungslehrkraft erworben hatten. 75% aller Lehramtsstudierenden der ersten Kohorte verfügten zum Zeitpunkt der ersten Untersuchung über mindestens einen Monat nebenberuflicher Lehrerfahrung. Dahingegen waren dies für die zweite Kohorte nur 21%.

Im Rahmen dieses Forschungsprojektes wurden vier Blockseminare durchgeführt. Alle Seminare fanden an drei ganzen, aufeinanderfolgenden Tagen statt. An jedem Tag wurde jeweils eine der folgenden, auf Borich (2015) zurückgehenden Dimensionen für effektive Klassenführung thematisiert: Raumordnung, Regeln und Routinen sowie Low-Profile (Borich, 2011). In allen vier Seminaren wurden die gleichen sieben Videoabschnitte analysiert. In den zwei Seminaren mit situiertem Ansatz wurden die Videos ohne theoretische und methodische Hinweise bearbeitet. Die Studierenden wurden dafür in mehreren Übungen dazu aufgefordert, Auffälligkeiten im Video zu explizieren, diese in Verbindung mit vorher gesammelten Praxiserfahrungen zu bringen sowie in Kleingruppen zu diskutieren. Im Gegensatz hierzu wurde die Videoanalyse in den zwei Seminaren mit kognitivistischem Ansatz immer durch einen Vortrag zur entsprechenden Klassenführungsstrategie eingeleitet. Die Übungen in diesen zwei Lehrveranstaltungen waren deutlich stärker strukturiert und

orientierten sich grob an Rating- und Analyseverfahren der quantitativen Unterrichtsanalyse (Appel & Rauin, 2016). Darüber hinaus wurden die zwei Gruppen, die an diesem Seminartyp teilnahmen, nicht explizit zur Reflexion auf die Praxiserfahrungen aufgefordert.

Für die Erfassung von Lerneffekten im Rahmen des vorliegenden Dissertationsprojekts wurde in Anlehnung an einen Videotest von König (König, 2015b) ein neuer Videotest namens CMC (Classroom Management Competence) entwickelt. CMC bestand aus vier kurzen Videovignetten mit insgesamt 30 größtenteils open-response Items. Die Items orientierten sich an den drei oben definierten Klassenführungsstrategien nach Borich (Borich, 2011) und bezogen zwei der drei situationsspezifischen Fähigkeiten, *perception* und *interpretation*, ein (König & Lebens, 2012). Testteilnehmer schauten sich ein Video auf einer zentral im Testraum positionierten Leinwand an und hatten acht Minuten Zeit die dazu gehörenden Fragen auf einem von der Testleitung bereitgestellten Tablet auszufüllen. Die Testangaben wurden mittels vorher definierter Kriterien kodiert, welche allesamt, richtige Antworten auf eine jeweilige Frage darstellten. Sie wurden über drei Runden von Expertenbefragungen validiert und ihre relative Relevanz bestimmt. Wenn ein Kriterium auf die von einer Testperson gegebene Antwort zutraf, wurde eine 1 für das entsprechende Kriterium vergeben. Der Testscore wurde durch die gewichtete Addition der Werte für sämtliche Kriterien gebildet.

Der Videotest wurde in einer kleinen Stichprobe (n = 29) pilotiert. Die Ergebnisse der Itemanalyse führten zur Eliminierung und neue Formulierung von vier Items sowie zur Änderung der Formulierung von acht weiteren Items, die offensichtlich zu schwierig oder zu einfach gewesen waren oder im Nachhinein als missverständlich formuliert eingeschätzt wurden. Außerdem wurde eine Kerngruppe von 18 guten Items definiert für die eine interne Konsistenz von α = .76 festgestellt wurde.

Neben der Videotest-Daten wurden auch die Daten von vier Items der abschließend im Rahmen aller vier Seminare durchgeführten Lehrveranstaltungsevaluation mitanalysiert (LUQ, 2018). Durch die Analyse dieser Sekundärerhebung wurden Unterschiede in der subjektiven Wahrnehmung der Studierenden erfasst (s. Tabelle 1).

Ergebnisse

Die Itemanalyse der Haupterhebung zeigte, dass die Werte zur internen Konsistenz deutlich niedriger als in den Richtlinien für Testentwicklung waren (Fisseni, 2004). Auch nach der

Eliminierung von 10 Items mit ungenügenden psychometrischen Werten, zeigte die Analyse der internen Konsistenz für den Pretest einen Wert von α = .57 und für den Posttest einen Wert von α = .56. Diese Werte deuten auf ein relativ hohes Maß an Messfehlern und eine eingeschränkte Testreliabilität hin. Dennoch konnte eine Validierungsuntersuchung die Hypothesen zur Konstruktvalidität bestätigen.

Die Analyse der Daten der Seminarevaluationen zeigte bei einigen Mittelwertvergleichen signifikante Unterschiede nach der Bonferroni-Korrektur der Alphafehlerkumulierung (Eid et al., 2015). So haben Lehramtsstudierende der ersten Kohorte mit Praxissemestererfahrung den Wissenszuwachs als deutlich höher empfunden als Lehramtsstudierende der zweiten Kohorte ohne Praxissemestererfahrung (F(1, 84) = 29.22, p < .001, d = 1.18).

Tabelle 1. Mittelwerte und Standardabweichungen zu vier Items der Seminarbeurteilungen

	1Sit.			1Cog.			2Sit.			2Cog.		
Item	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
1. Der Besuch der Veranstaltung führt zu einem spürbaren Wissenszuwachs.	24	3,7	1,1	20	4,1	1,3	20	4,9	1,2	22	5,4	0,8
2. Das Tempo der Veranstaltung ist angemessen.	24	4,7	1,8	20	4,3	1,4	20	5,3	1,2	22	5,1	1,3
3. Der in der Veranstaltung vermittelte Stoff ist gut strukturiert.	24	4,6	1,1	20	5,2	0,8	20	5,3	1,0	22	5,3	1,2
4. In der Veranstaltung werden Medien in für den Inhalt geeigneter Weise eingesetzt.	24	5,5	0,9	20	5,7	0,6	20	5,7	0,6	22	5,6	1,1

Anm. 1Sit. = Seminargruppe aus der ersten Kohorte mit situiertem Ansatz; 1Cog. = Seminargruppe aus der ersten Kohorte mit kognitivistischem Ansatz; 2Sit. = Seminargruppe aus der zweiten Kohorte mit situiertem Ansatz; 2Cog. = Seminargruppe aus der zweiten Kohorte mit kognitivistischem Ansatz.

Bzgl. der Videotestwerte zeigten die Studierenden der ersten Kohorte mit Praxissemestererfahrung beim Pretest höhere Werte (M = 23,1, SE = 0,79) für Klassenführungskompetenz als die Studierenden der zweiten Kohorte ohne Praxissemestererfahrung (M = 21,8, SE = 0,68). Dieser Unterschied war allerdings nicht signifikant (t(85) = 1,16, p = ,25). Die Auswertung der Mittelwertvergleiche des selbstentwickelten Videotests zeigen einen großen signifikanten Haupteffekt auf Testebene für alle Seminargruppen zwischen Pretest und Posttest (F(1, 83)= 27,14; p= ,000; η^2= .246) an. Dies bedeutet, dass die beteiligten Lehramtsstudierenden im Durchschnitt deutliche Kompetenzgewinne vorzeigen konnten (s. Tabelle 2).

Tabelle 2. Mittelwerte und Standardabweichungen für jede Seminargruppe in Pretest und Posttest für CMC

Variable	1Sit.[a]		1Cog.[b]		2Sit.[c]		2Cog.[d]	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Pretest	22,1	5,0	24,2	5.3	22,0	5,1	21,7	3,9
Posttest	24,8	5,0	26,0	5.0	25,6	3,4	25,5	4,7

Anm. 1Sit. = Seminargruppe aus der ersten Kohorte mit situiertem Ansatz; 1Cog. = Seminargruppe aus der ersten Kohorte mit kognitivistischem Ansatz; 2Sit. = Seminargruppe aus der zweiten Kohorte mit situiertem Ansatz; 2Cog. = Seminargruppe aus der zweiten Kohorte mit kognitivistischem Ansatz.
[a]*n* = 24. [b]*n* = 20. [c]*n* = 20. [d]*n* = 23.

Beide Interaktionseffekte erster Ordnung zwischen Testscore und Praxiserfahrung ($F(1, 83) = 1{,}69$, $p = ,20$, $\eta^2 = ,020$) und zwischen Testscore und Seminartyp ($F(1, 83) = 0{,}10$, $p = ,75$, $\eta^2 = .001$) zeigen keine signifikanten Effekte. Wider Erwarten konnten aber auch keine signifikanten Interaktionseffekte zweiter Ordnung zwischen Testscore, Praxiserfahrung und Seminartyp festgestellt werden ($F(1, 83) = 0{,}27$, $p = ,61$, $\eta^2 = .003$ und $F(1, 83) = 0{,}27$, $p = ,61$, $\eta^2 = ,003$).

Diskussion

Die Ergebnisse dieser Studie zeigen erstens, dass im Durchschnitt alle Lehramtsstudierenden, die an der Studie teilgenommen haben, ihre Klassenführungskompetenz deutlich steigern konnten. Darüber hinaus zeigt sich allerdings, dass es im Rahmen dieser Intervention keine signifikanten Unterschiede in der Kompetenzänderung im Vergleich von Studierenden mit und ohne Praxissemestererfahrungen sowie hinsichtlich des unterschiedlichen didaktischen Designs der videobasierten Seminare gegeben hat. Damit werden frühere Forschungsbefunde bestätigt, dass für die universitäre Lehrerbildung beide didaktischen Ansätze eingesetzt werden können (Blomberg et al., 2014; Kumschick et al., 2017; Syring et al., 2015). Allerdings bestätigen diese Ergebnisse nicht die Annahme, dass Lehramtsstudierende in der ersten Phase ihres Studiums von Videovignetten schnell kognitiv überfordert seien. Darüber hinaus zeigen diese Ergebnisse, dass die Verfügbarkeit von Praxiserfahrungen für die Vermittlung von Klassenführungskompetenz keine entscheidende Rolle spielt. Dennoch lässt der große Unterschied bezüglich des subjektiv wahrgenommenen Wissenszuwachses vermuten, dass es für das Thema Klassenführung relevante Unterschiede zwischen den Studierenden mit und ohne Praxissemestererfahrungen gibt, die im Rahmen dieser Studie nicht näher untersuchte werden konnten. Diese Vermutung wird durch die Mittelwertunterschiede beim Videotest im Pretest und bei den Angaben hinsichtlich des empfundenen Tempos der Veranstaltung

unterstützt. Obwohl beide Unterschiede nicht signifikant sind, lässt die Richtung, Effektgröße und die Höhe des p-Wertes vermuten, dass ein genaueres Messinstrument möglicherweise signifikante Unterschiede hätte finden können beziehungsweise dass das Tempo des Kurses von Lehramtsstudierenden ohne Praxissemestererfahrung als deutlich angemessener empfunden wurde.

Die wichtigste Einschränkung dieser Studie liegt in der limitierten Reliabilität des Instrumentes. Diese Werte deuten auf relativ ungenaue Messungen hin, was die Findung von signifikanten Effekten deutlich erschwert. Darüber hinaus zeigte sich, dass Lehramtsstudierende der ersten Kohorte mit Erfahrung im Praxissemester ihr Seminar deutlich weniger gewinnbringend beurteilt haben. Da das Seminar in der ersten Kohorte zum ersten Mal durchgeführt wurde und bei der zweiten Kohorte die Erfahrungen aus dem ersten Durchgang eine Rolle gespielt haben könnten, sind Übungseffekte bei den Dozenten nicht auszuschließen. Da allerdings alle Vorträge, Übungen, Abläufe und Medien für beide Kohorten standardisiert wurden, dürften etwaige Übungseffekte gegebenenfalls nur einen geringen Einfluss gehabt haben. Es ist dadurch wahrscheinlich, dass Studierende der ersten Kohorte das Seminar als deutlich einfacher wahrgenommen haben als Studierende der zweiten Kohorte.

Da diese Studie Indizien dafür liefert, dass Praxiserfahrungen die Klassenführungskompetenz von Lehramtsstudierenden beeinflussen, könnten zukünftige Studien sich weiter auf diese Frage fokussieren. Dabei wären sicherlich auch nicht-kognitive Kompetenzaspekte wie zum Beispiel Überzeugungen zu Klassenführungsstils (Martin et al., 2007) zu betrachten. Darüber hinaus könnten zukünftige Studien über eine Längsschnitterhebung der Frage nachgehen, ob für die langfristige Entwicklung von Klassenführungskompetenz ein Praxissemester als Basis für darauffolgende universitäre Reflexion belegt werden sollte, oder ob ein videobasiertes Seminar zu Klassenführung vorzugsweise als Vorbereitung zum Praktikum belegt werden sollte.